JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 5 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Letters to the Editor

Viewpoint

Noninvasive Bioimpedance Methods From the Viewpoint of Remote Monitoring in Heart Failure

Pawel Krzesinski[1], MD, PhD; Aleksander Sobotnicki[2], MSc, PhD; Adam Gacek[2], MSc, PhD; Janusz Siebert[3], MD, PhD; Andrzej Walczak[4], MSc, PhD; Piotr Murawski[5], MSc, PhD; Grzegorz Gielerak[1], MD, PhD

[1]Department of Cardiology and Internal Diseases, Military Institute of Medicine, Warsaw, Poland

[2]Institute of Medical Technology and Equipment (ŁUKASIEWICZ-ITAM), Zabrze, Poland

[3]Department of Family Medicine, Medical University of Gdansk, Gdansk, Poland

[4]Software Engineering Department, Cybernetics Faculty, Military University of Technology, Warsaw, Poland

[5]Department of Informatics, Military Institute of Medicine, Warsaw, Poland

Corresponding Author:
Pawel Krzesinski, MD, PhD
Department of Cardiology and Internal Diseases
Military Institute of Medicine
Szaserow 128
Warsaw, 04-141
Poland
Phone: 48 606939390
Email: pkrzesinski@wim.mil.pl

Abstract

Heart failure (HF) is a major clinical, social, and economic problem. In view of the important role of fluid overload in the pathogenesis of HF exacerbation, early detection of fluid retention is of key importance in preventing emergency admissions for this reason. However, tools for monitoring volume status that could be widely used in the home setting are still missing. The physical properties of human tissues allow for the use of impedance-based noninvasive methods, whose different modifications are studied in patients with HF for the assessment of body hydration. The aim of this paper is to present the current state of knowledge on the possible applications of these methods for remote (home-based) monitoring of patients with HF.

(JMIR Mhealth Uhealth 2021;9(5):e25937) doi:10.2196/25937

KEYWORDS

heart failure; impedance cardiography; remote monitoring; overhydration; hemodynamics; heart; cardiac function; cardiac; monitor

Introduction

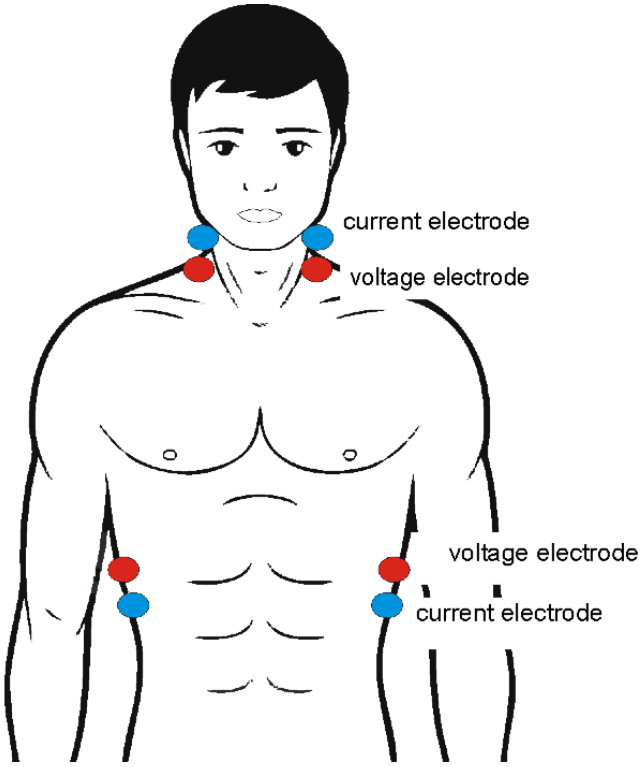
Heart failure (HF) is a major clinical, social, and economic problem, which is attributable, among other factors, to the high frequency of exacerbations requiring urgent hospital admission. The 30-day and 6-month rates of decompensated HF are 19% to 31% and up to 50%, respectively [1-3]. In view of the important role of fluid overload in the pathogenesis of HF exacerbation, early detection of fluid retention is of key importance for preventing emergency admissions. This is theoretically possible, as there is a time window of 10 to 20 days between the onset of fluid overload and hospital admission due to decompensated HF [4]. However, tools for monitoring volume status that could be widely used in the home setting are still missing. The complexity of the pathophysiological processes involved in fluid retention and redistribution hinders the possibility of predicting exacerbation episodes with adequate sensitivity and specificity based solely on body weight monitoring [5]. This is desirable for the assessment of not only the overall fluid status, but also fluid levels in different body compartments, with a particular focus on the chest. The possible assessment of hemodynamic parameters, such as heart rate, stroke volume (SV), and cardiac output (CO), has important added value in the assessment of the etiology of cardiovascular decompensation.

The physical properties of human tissues allow for the use of impedance-based (resistance-based) noninvasive methods, whose different modifications are studied in patients with HF, for the assessment of body composition. The aim of this paper is to present the current state of knowledge on the possible applications of these methods for remote (home-based) monitoring of patients with HF.

Impedance Cardiography

Impedance cardiography (ICG) is a noninvasive method designed for monitoring hemodynamic parameters on the basis of analysis of thoracic electrical resistance. Eight electrodes are symmetrically placed on both sides of the neck and along the mid-axillary line (Figure 1). Electrodes supplying current (so-called current electrodes) are placed on the neck above the voltage electrodes and on the chest below the voltage electrodes. Voltage electrodes receive the potential from the thoracic region, including, among others, the heart and large vessels in the field of electric current flowing between the current electrodes. In this way, it is possible to determine thoracic fluid content [6]. During the examination, voltage changes associated with changes in blood volume and velocity in large vessels during systole and diastole are also analyzed [6]. This enables the calculation of parameters, including SV and CO [6], which is a particular advantage of the method. The estimation of SV and CO is far more difficult than the evaluation of volume status, as it requires recording reliable signals of changes in impedance due to blood flow in large vessels. The need to refer this signal to an electrocardiographic curve poses a considerable technical challenge for devices used to assess the pumping function of the heart.

Figure 1. Electrode placement for tetrapolar impedance cardiography measurement.



ICG is also suitable for calculating parameters, including the Heather index, velocity index, and acceleration index, which are associated with the dynamics of left ventricular ejection [6]. Currently, several commercial stationary ICG devices are available, for example, Niccomo/Cardioscreen (Medis), Task Force Monitor (CNSystem), Hotman System (Hemo Sapiens Medical Inc), BioNex Impedance Cardiograph (Mindware Technologies Ltd), BioZ (CardioDynamics), NICO 100C (BIOPAC Systems), IQ2 (Noninvasive Medical Technologies Inc), and NICOM (Cheetah Medical, Inc).

The concepts of portable impedance rheographs were first proposed by researchers from the Warsaw University of Technology already more than 20 years ago. Cybulski et al [7] developed a Holter-type device prototype (ReoMonitor), which provides continuous recording of a cardiac impedance signal with simultaneous calculation of SV, left ventricular ejection time, and pre-ejection period. A comparison with echocardiography showed a high level of SV measurement consistency between these two methods (r=0.83). It was noted, though, that the quality of daytime recordings was lower (the percentage of properly recognized cardiac cycles was 20%-80%) compared to night-time recordings (75%-90%), which is most likely attributable to body movement and speech. Weyer et al [8] presented their concept of a mobile electrocardiographic system with multifrequency ICG assessment. The device has a wireless connection via Bluetooth, and uses a four-electrode (tetrapolar) topology identical to the one used in stationary devices. The measurements performed by the device were sent to an external module (a desktop computer). The measurements were compared with a commercial stationary device (Niccomo, Medis). A significant correlation between impedance (Z) and its time derivative (dZ/dt) was observed. The measurement of impedance resulted in a relative error below 1%, and on that basis, it was concluded that the device might be implemented in broader practice. As a possible useful solution, the application of textile electrodes for better wearing comfort was suggested. It was also noted that an active Bluetooth connection significantly increased power consumption (from 90 mW to 220 mW). To eliminate this problem, it was proposed that data transmission should not be continuous, but instead data should

be stored on a memory card and sent during periods between registrations [8].

Ulbrich et al [9] developed the IMPACT Shirt (IMPedAnce Cardiography Textile) designed for SV measurement. The examination is performed by a small device recording ICG and electrocardiography (ECG) curves. The signal is transmitted via Bluetooth to a smartphone or a computer. To date, proof-of-concept studies have shown high compatibility of measurements recorded using the IMPACT Shirt and commercially available devices.

Yazdanian et al [10] also developed a concept of a portable ICG system for SV assessment, with preliminary research showing significant correlations between SV measured using the cardiac impedance method and SV determined by Doppler echocardiography (r=0.89, P<.05). Panfili et al [11] presented a device for the continuous monitoring of CO, with transmission of the recorded signal via Bluetooth. Other researchers also tended to implement ICG devices for remote monitoring to a greater or lesser extent in their scientific concepts [12-15].

Recently, a team of researchers working on the AMULET project [16] developed a concept for remote ICG measurements with automatic data transfer to a remote electronic platform (Figure 2). Measurements are performed by a miniaturized ICG recorder measuring biological signal impedance curves and ECG, which are sent to a smartphone and then to an electronic platform, where the recorded signals are analyzed. An additional mobile app supports the reporting of symptoms experienced by the patient at the time of measurement and other vital parameters (such as blood pressure and body mass). All data are integrated and included in the patient's individual record. There is also an option to preview recorded curves and track the trend of changes in thoracic fluid content over time (Figure 3). In addition, the system allows for sending medical recommendations based on the interpretation of recorded data to the patient. Currently, there is an ongoing pilot study assessing the feasibility of measurements and the consistency of results with a reference stationary device.

Figure 2. Concept of remote impedance cardiography measurements in the AMULET project.

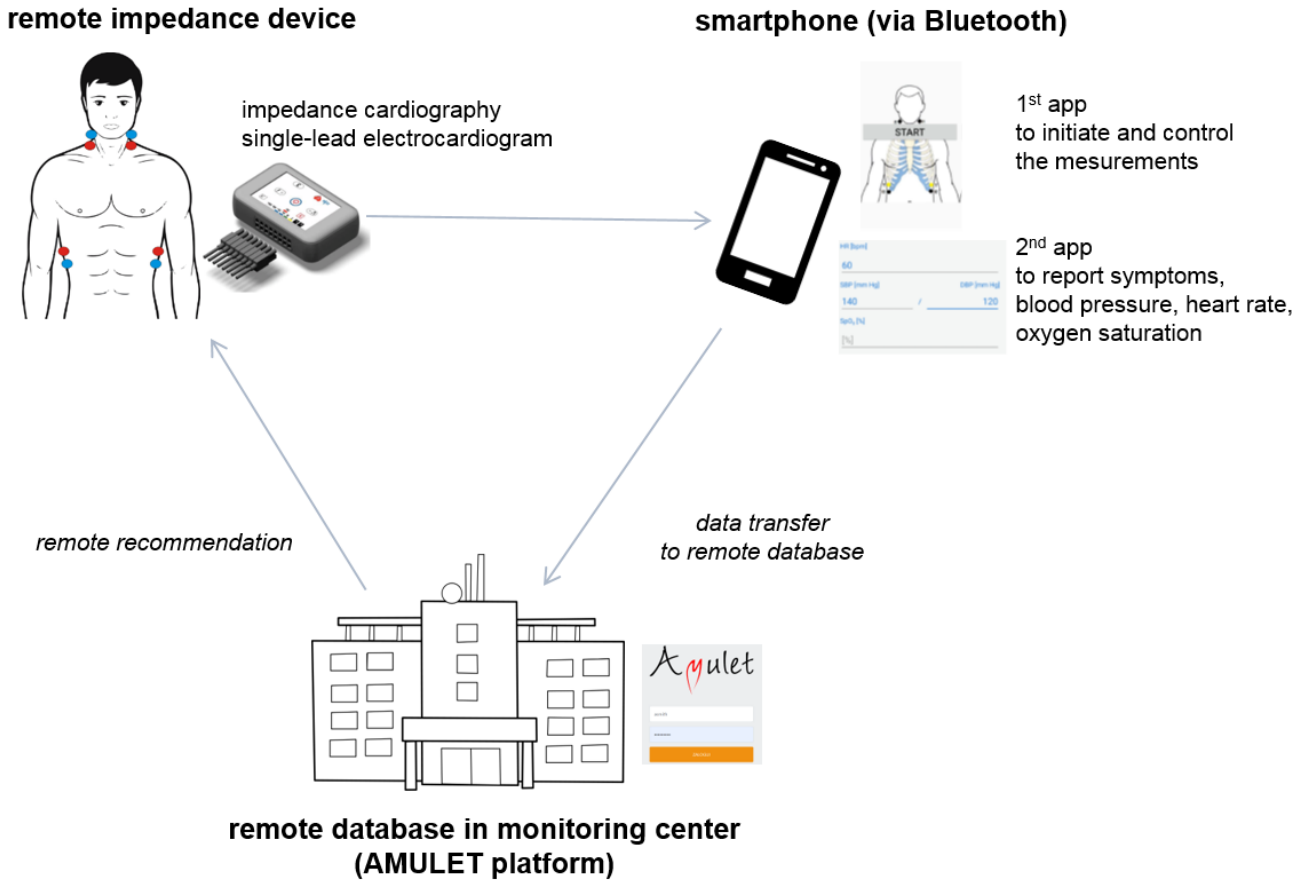
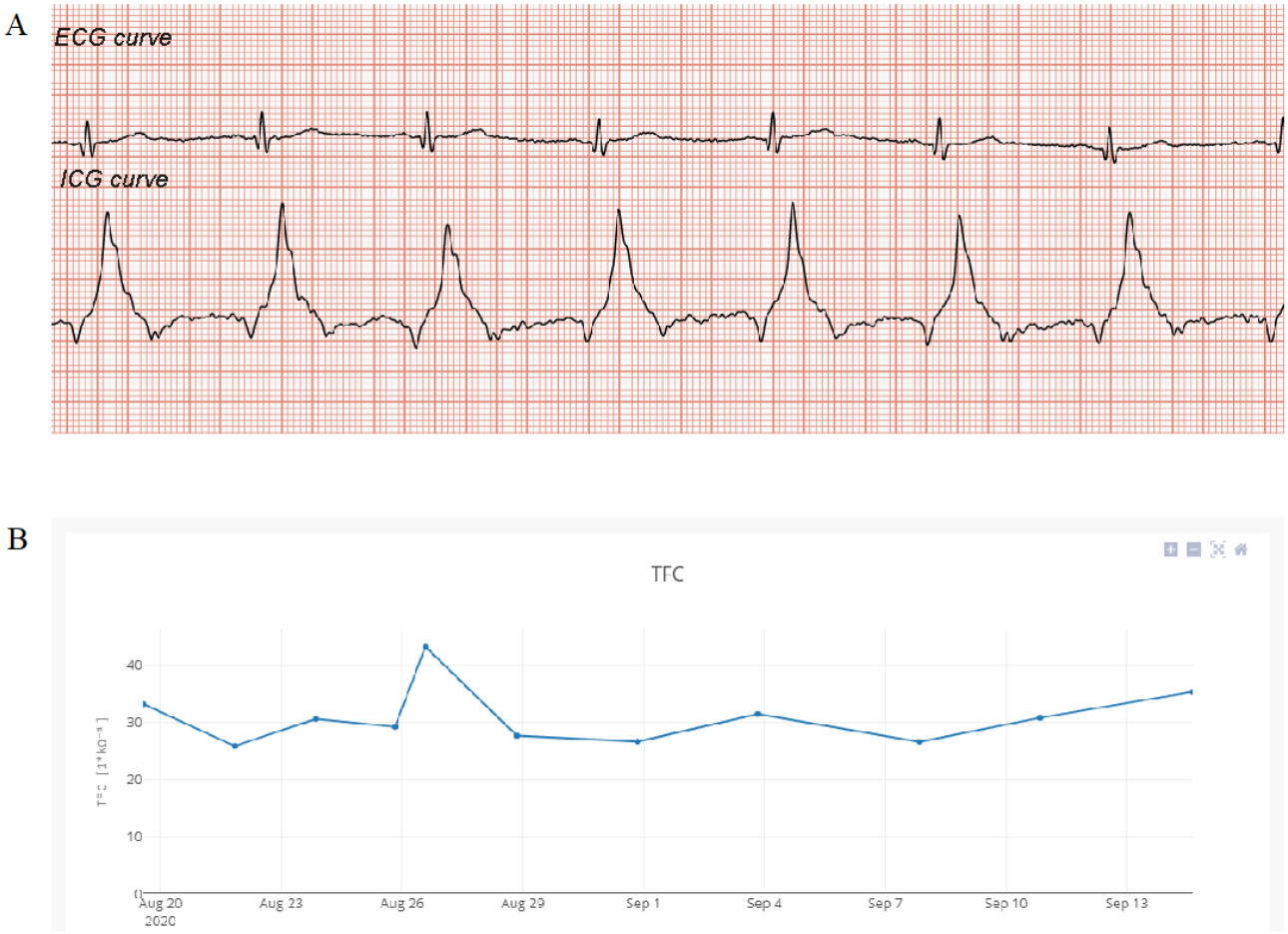


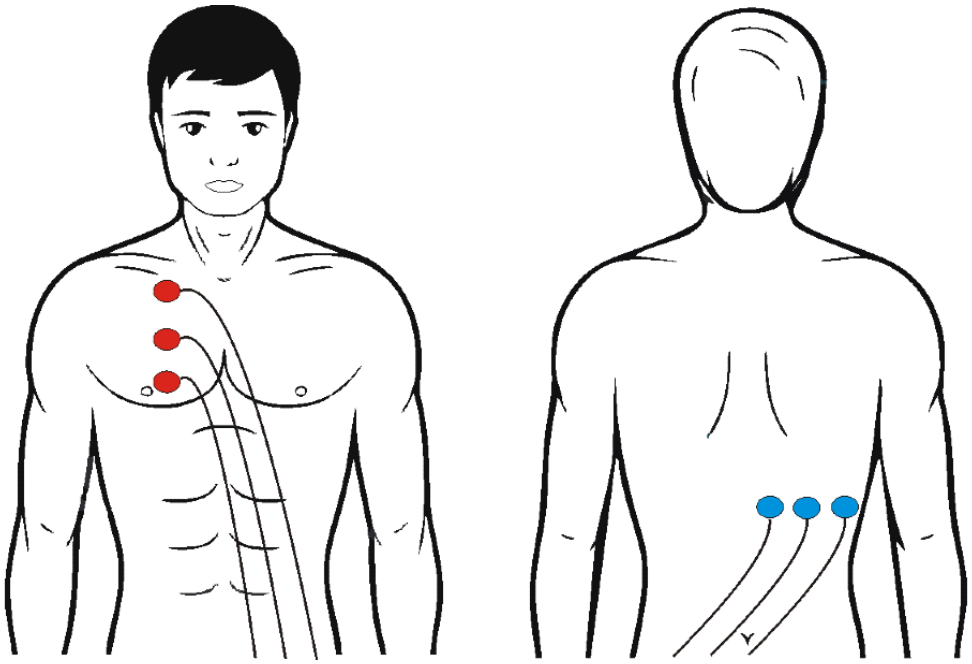
Figure 3. Remote database screenshots. (A) Recorded electrocardiography (ECG) and impedance cardiography (ICG) curves (the first time derivative of thoracic electrical impedance [dZ/dt]). (B) The trend of changes in thoracic fluid content (TFC) over time.



Lung Impedance

A measurement system based on the impedance method, using modified electrode placement, was also developed (lung impedance [LI]/internal thoracic impedance [ITI]). Three electrodes are placed vertically on the front upper right side of the chest, and another three electrodes are placed on the back along the horizontal line, below the right scapula (Figure 4).

Figure 4. Electrode placement for lung impedance measurement.



This modification is intended to eliminate the impact of large thoracic vessels on the base impedance, which is referred to as LI in this case. It was found that classical ICG measurements in the long axis were associated with a significant contribution of blood flow to chest impedance analysis, since blood conductance is 10 to 12 times higher than that of lung tissues. According to the authors' assumptions, the LI method eliminates some of these limitations, which increases its sensitivity to changes in lung fluid content. However, it should be emphasized that the applied electrode topography limits the assessment of lower lung segments, where fluid accumulation is most likely to occur first.

In 2006, a study was published [17] to report on the use of the Edema Guard Monitor (EDM, model RS-207, RS Medical Monitoring) for the assessment of the utility of ITI in predicting cardiac decompensation in patients hospitalized for cardiac reasons (acute coronary syndrome, valvular disease, or HF). The relationship between reduced ITI and the risk of pulmonary edema (confirmed clinically and radiologically) was assessed in a total of 265 patients (aged between 37 and 83 years). ITI from the first measurement was considered baseline and was used as a reference for further measurements performed every 30 minutes (for up to 24 hours or until pulmonary edema). For the purpose of the analysis, the patients were classified into the following two groups: those with pulmonary edema (DE, n=37) and those without (DNDE, n=228). ITI fluctuations in the DNDE group fell within the range of +11.0% to −10.1%. ITI deceased by over 12% in all patients in the DE group at 30 minutes (26 patients) and up to 60 minutes (11 patients) before the onset of signs of pulmonary edema. The patients presented with no symptoms of pulmonary edema and had a stable blood oxygen saturation and respiratory rate on reaching this threshold. It was concluded that a change in ITI of no more than 10% essentially excluded the risk of pulmonary edema.

These experiences were used in another study conducted in patients with ST-elevated myocardial infarction (STEMI) [18]. A hypothesis was formulated that LI-guided therapy initiated at the asymptomatic stage of pulmonary edema (Killip class I) would prevent cardiovascular destabilization and thus improve patient prognosis. The study included 560 patients with STEMI and no signs of HF on admission. During their hospitalization, repeated physical examinations, including blood pressure, heart rate, and oxygen saturation measurements, were performed. LI was monitored every 60 minutes using an impedance device (RSMM Company). Based on a pilot study, the authors selected an ITI fluctuation range from 12% to 14% as an indicator of the transition from interstitial to alveolar edema (sensitivity 96%-100%, specificity 100%). As a result, patients with LI drop of not more than 12% were included in group 1. Others were randomized to group 2 (controls) and group 3 (intervention) at a 2:1 ratio. Diuretic treatment with furosemide (oral followed by intravenous) was immediately initiated and maintained until reaching baseline LI in the intervention group. Physicians treating the control group were blinded to LI values. After discharge, the patients were followed for 6 months for the assessment of events. Of the total of 560 study patients, 347 (62%) showed a change in LI of 12% or less. None reported dyspnea, and all of them were still classified as Killip I. Of the 213 patients with LI over 12%, 142 were included in the control group (group 2) and 71 were assigned to the LI-guided therapy group (group 3). The following events were observed in groups 2 and 3: a total of 132 patients halted at Killip class IIC/III, and 10 patients presented with Killip class IV symptoms (including six deaths). In group 3, only 11% of patients developed pulmonary edema (no Killip class IV, no deaths), whereas all patients in group 2 developed pulmonary edema, including six deaths (P<.001). The 1-year readmission rate, 6-year mortality rate, and new-onset HF rate were lower in group 3 than in group 2.

These findings became the starting point for applying the method in outpatient care. A prospective, open-label, single-center study was conducted in a group of 222 patients with HF and left ventricular ejection fraction (LVEF) less than 35%, who were hospitalized within 12 months of recruitment with HF exacerbation [19]. Stable patients whose LI was considered reference (basal LI) were recruited. A total of 388 episodes of HF exacerbation were reported during long-term follow-up (mean 32 months). The episodes were preceded by LI decrease 2 to 4 weeks before symptom onset. One-year risk of decompensation for a 30% to 40% drop in LI was 42-fold higher compared to that for a 20% or less drop in LI, and it was more than 130 times higher for an over 40% drop. In 2016, a prospective interventional study was conducted to assess the effects of LI-guided outpatient treatment in preventing patient readmission with HF exacerbation (IMPEDANCE-HF Trial, ClinicalTrials.gov NCT01315223) [20]. A group of 252 patients after an episode of HF exacerbation (LVEF ≤35%, New York Heart Association [NYHA] II-IV) were randomized to the LI-guided treatment arm or the control arm. After 48 months (SD 32) of follow-up in the intervention group and 39 months (SD 26) of follow-up in the control group, it was found that LI-guided therapy was associated with, among other things, reduced rates of readmission caused by HF exacerbation (annual rate 0.41 vs 0.94; hazard ratio [HR] 0.63, 95% CI 0.53-0.75; P<.001; number needed to treat [NNT] 1.9), reduced all-cause mortality (annual rate 0.08 vs 0.14; HR 0.52, 95% CI 0.35-0.78; P=.002; NNT 7.5), reduced CV mortality (annual rate 0.05 vs 0.11; HR 0.41, 95% CI 0.25-0.67; P<.001; NNT 6.1), and reduced HF mortality (annual rate 0.03 vs 0.08; HR 0.35, 95% CI 0.15-0.58; P=.001; NNT 7.1). An analysis of changes in LI (ΔLIR) showed reduced ΔLIR 3 weeks prior to admission.

Spectroscopy: Bioelectrical Impedance Analysis

Bioelectrical impedance analysis (BIA), also referred to as bioimpedance analysis, is based on the measurement of impedance (ie, electrical resistance consisting of two elements: resistance and reactance) of tissues through which low amperage electric current (≤1 mA) is passed. The phenomenon of resistance is associated with the specific resistance of different tissues, while reactance is associated with the electrical capacity of cell membranes. BIA is a method most commonly used for estimating body composition.

A group of French researchers from Institute des Nanotechnologies de Lyon developed a device for home-based

wireless segmental BIA measurement. Although patients with end-stage renal disease were primarily indicated as the potential target group, the proposed solution may prove beneficial in all patients requiring regular and objective monitoring of body fluid content [21]. The device may connect to a smartphone featured with a mobile app via Bluetooth [22]. Preliminary recordings have shown that the proposed solution is sensitive to dialysis-related changes in impedance for all body compartments. However, a significant influence of body position on the results of segmental BIA has been revealed, which may lead to measurement misinterpretation, for example, during classical hemodialysis [23]. Impedance at the electrode-skin interface plays a much more important role in segmental measurements than whole-body BIA measurement. These limitations should be duly considered in the optimization of device design.

Bioimpedance spectroscopy [24], a method related to multifrequency bioimpedance, is used for home-based assessment of thoracic fluid status. It may be complementary to other approaches, for example, natriuretic peptide measurement and clinical assessment. When used in this combination, it helps to identify patients with fluid overload and worse prognosis [25]. The method enables a snap-shot home-based assessment with rapid transfer of data to the monitoring center. In their pilot study, Beckmann et al [26] showed that spectroscopic assessment (Xitron Hydra 4200, Xitron Technologies) was a valuable tool for estimating changes in body fluid volumes. When using the method to monitor five patients on diuretic therapy due to pulmonary edema, it was found that the overall assessment of whole-body and thoracic impedance allowed for reliable tracking of fluid redistribution and dehydration.

Dovancescu et al [27] developed a portable monitoring system based on a wearable fluid accumulation vest with an option to send data to a telemonitoring center via a smartphone. The vest is equipped with four textile electrodes placed pairwise on either side of the rib cage. An electronic module is placed on the back and communicates with the smartphone via Bluetooth. The system also comprises a smartphone app and an electronic database. The device enables assessment of intra- and extracellular fluid, respiratory rhythm, one-lead ECG, and patient posture and motion. A single measurement takes approximately 10 minutes, with automatic data transfer to an electronic database after completion.

An assessment of the system for predicting HF exacerbation in patients discharged after hospitalization for decompensation was planned in the SENTINEL-HF study (ClinicalTrials.gov NCT01877369) [27]. Enrollment of 180 NYHA II-IV patients hospitalized with HF in two different centers was scheduled, with patient management including daily assessment of thoracic fluid status, home visits (on days 7 and 45), telephone contact (day 14), and follow-up at endpoints (90 days). The study was launched in 2013, with the last patient recruited in April 2015. Unplanned HF-related rehospitalization along with HF exacerbation requiring increased doses of diuretics and/or associated with life quality deterioration was adopted as a composite endpoint. Decompensation alerts are triggered in the studied system based on an algorithm that tracks changes in volume status and compares them with normal variability in a given patient.

Preliminary findings from the study were presented by Darling et al [28]. The ability of patients to perform the procedures indicated in the SENTINEL-HF study and the predictive value of bioimpedance for predicting episodes of unplanned hospitalization for HF or the need to increase the dose of diuretic were assessed. Of 180 patients who consented to participate in the study, 106 completed the 75-day follow-up period; however, only 57 (53.8%) provided bioimpedance data of adequate quality. Technical factors, including noise artefacts and app malfunction, were the main reasons. Despite a relatively high predictive value of the algorithm (87% sensitivity, 70% specificity), the above-described data loss led the authors to conclude that further studies were needed to improve patient compliance and eliminate technical limitations. It should be emphasized that the device assessed in the SENTINEL study has not yet been approved by the Food and Drug Administration (FDA), and the final results of SENTINEL-HF have not been released yet.

The concept of a vest with four textile electrodes for the measurement of teletransmission was presented by Cuba-Gyllensten et al [29]. The electrodes are arranged pairwise on an elastic belt, ensuring fixation at a constant distance. The study used a prototype of the device, performing measurements at 60 different frequencies (10 kHz to 1 MHz). A number of measurements were carried out in 20 patients hospitalized with acute decompensated HF (mean age 74.7 years, mean LVEF 37%) and then correlated with clinical indicators. A significant relationship was found between changes in impedance and changes in body weight (r=−0.830, P<.001) and HF severity score (r=−0.537, P<.001). Correlations were also identified between bioimpedance measurements and LVEF (r=0.450, P=.047) and N-terminal pro-brain natriuretic peptide (r=−0.41, P=.038). The vital status of the subjects was assessed at 18 months after discharge. Three of four patients presenting with low impedance at discharge (<20 Ω) died of HF within the follow-up period. Only one death was noted among 16 other patients who left the hospital with impedance exceeding 20 Ω. The impedance values significantly differed between subjects who died and survivors, both at admission (P=.003) and at discharge (P=.02) [30].

In 2015, Seulki et al [31] presented a new wearable bioimpedance device to assess pulmonary fluid status using electrodes placed in the lower thoracic region. The idea was developed to reduce interference with other clinical equipment that is usually placed on the neck and thorax. In this system, four electrodes (two voltage and two current) are attached on the skin over the lower left thoracic region in a 2×2 array configuration with 5 cm and 12 cm center-to-center distances for the horizontal and vertical directions, respectively. Using this approach in the monitoring of patients admitted to hospital with HF decompensation, the authors revealed a good correlation of impedance changes with fluid loss (R^2>0.80) [31]. The prognostic value of in-hospital monitoring with this wearable bioimpedance device was evaluated in a prospective study involving 36 patients (mean age 81 years, mean LVEF 45%)

with acute decompensated HF [32]. At 1-year follow-up after discharge, only 12% (3/24) of patients with an increase in impedance within hospitalization died compared with 50% (6/12) of patients with an impedance decrease (*P*=.01). Although the difference for HF readmission did not reach statistical significance (17% vs 33%, *P*=.28), it was unequivocal (25% vs 75%, *P*=.01) for the composite endpoint (HF hospitalization and all-cause mortality), with a HR of 4.96 (*P*=.003) [32].

Conclusions

Although most of the diagnostic methods discussed above remain in the stage of scientific research, they have a great potential for implementation with the use of all the advantages of telemedicine. On the other hand, multiple concerns have been raised regarding the ability of elderly patients to perform measurements at home. It turns out, however, that this factor is not necessarily a major limitation of remote monitoring. Aamodt et al [33] showed that most patients (>80%) with advanced HF were able to correctly perform home impedance measurements and send back electronic symptom questionnaires. Wireless and portable sensors, which are particularly useful for home-based monitoring, may also be an important step paving the way for these technologies [34].

The use of devices for the assessment of the volume status and hemodynamic profile in a noninvasive and patient-friendly manner should be one of the main directions in the development of telecardiology. These devices may enable systematic and frequent monitoring, which would represent a significant contribution to the complex system of care for patients with HF.

Acknowledgments

This work was supported by the National Centre for Research and Development and prepared within the framework of the scientific project “A new Model of medical care with Use of modern methods of non-invasive cLinical assEssment and Telemedicine in patients with heart failure” (STRATEGMED3/305274/8/NCBR/2017).

Conflicts of Interest

All authors are executives of the project STRATEGMED3/305274/8/NCBR/2017.

References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, 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]
2. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 02;360(14):1418-1428. [doi: 10.1056/NEJMsa0803563] [Medline: 19339721]
3. Epstein AM, Jha AK, Orav EJ. The relationship between hospital admission rates and rehospitalizations. N Engl J Med 2011 Dec 15;365(24):2287-2295. [doi: 10.1056/NEJMsa1101942] [Medline: 22168643]
4. Schiff GD, Fung S, Speroff T, McNutt RA. Decompensated heart failure: symptoms, patterns of onset, and contributing factors. Am J Med 2003 Jun 01;114(8):625-630. [doi: 10.1016/s0002-9343(03)00132-3] [Medline: 12798449]
5. Desai AS. Home monitoring heart failure care does not improve patient outcomes: looking beyond telephone-based disease management. Circulation 2012 Feb 14;125(6):828-836 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.031179] [Medline: 22331920]
6. Krzesiński P, Gielerak G, Kowal J. [Impedance cardiography - a modern tool for monitoring therapy of cardiovascular diseases]. Kardiol Pol 2009 Jan;67(1):65-71. [Medline: 19253194]
7. Cybulski G, Koźluk E, Michalak E, Niewiadomski W, Piatkowska A. Holter-type impedance cardiography device. A system for continuous and non-invasive monitoring of cardiac haemodynamics. Kardiol Pol 2004 Aug;61(8):138-146. [Medline: 15457280]
8. Weyer S, Menden T, Leicht L, Leonhardt S, Wartzek T. Development of a wearable multi-frequency impedance cardiography device. J Med Eng Technol 2015 Feb;39(2):131-137. [doi: 10.3109/03091902.2014.990161] [Medline: 25559781]
9. Ulbrich M, Mühlsteff J, Sipilä A, Kamppi M, Koskela A, Myry M, et al. The IMPACT shirt: textile integrated and portable impedance cardiography. Physiol Meas 2014 Jun;35(6):1181-1196. [doi: 10.1088/0967-3334/35/6/1181] [Medline: 24846072]
10. Yazdanian H, Mahnam A, Edrisi M, Esfahani MA. Design and Implementation of a Portable Impedance Cardiography System for Noninvasive Stroke Volume Monitoring. J Med Signals Sens 2016;6(1):47-56 [FREE Full text] [Medline: 27014612]
11. Panfili G, Piccini L, Maggi L, Parini S, Andreoni G. A wearable device for continuous monitoring of heart mechanical function based on impedance cardiography. Conf Proc IEEE Eng Med Biol Soc 2006;2006:5968-5971. [doi: 10.1109/IEMBS.2006.260250] [Medline: 17946352]

12. Ferreira J, Seoane F, Lindecrantz K. Portable bioimpedance monitor evaluation for continuous impedance measurements. Towards wearable plethysmography applications. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:559-562. [doi: 10.1109/EMBC.2013.6609561] [Medline: 24109748]
13. Nakagawara M, Yamakoshi K. A portable instrument for non-invasive monitoring of beat-by-beat cardiovascular haemodynamic parameters based on the volume-compensation and electrical-admittance method. Med Biol Eng Comput 2000 Jan;38(1):17-25. [doi: 10.1007/BF02344683] [Medline: 10829385]
14. Willemsen GH, De Geus EJ, Klaver CH, Van Doornen LJ, Carroll D. Ambulatory monitoring of the impedance cardiogram. Psychophysiology 1996 Mar;33(2):184-193. [doi: 10.1111/j.1469-8986.1996.tb02122.x] [Medline: 8851246]
15. Pinheiro E, Postolache O, Girão P. Contactless Impedance Cardiography Using Embedded Sensors. Measurement Science Review 2013;13(3):157-164. [doi: 10.2478/msr-2013-0025]
16. AMULET Project. URL: https://amulet.wim.mil.pl/english [accessed 2021-03-01]
17. Shochat M, Charach G, Meyler S, Meisel S, Weintraub M, Mengeritsky G, et al. Prediction of cardiogenic pulmonary edema onset by monitoring right lung impedance. Intensive Care Med 2006 Aug;32(8):1214-1221. [doi: 10.1007/s00134-006-0237-z] [Medline: 16775717]
18. Shochat M, Shotan A, Blondheim DS, Kazatsker M, Dahan I, Asif A, et al. Usefulness of lung impedance-guided pre-emptive therapy to prevent pulmonary edema during ST-elevation myocardial infarction and to improve long-term outcomes. Am J Cardiol 2012 Jul 15;110(2):190-196. [doi: 10.1016/j.amjcard.2012.03.009] [Medline: 22482863]
19. Shochat M, Shotan A, Blondheim DS, Kazatsker M, Dahan I, Asif A, et al. Derivation of baseline lung impedance in chronic heart failure patients: use for monitoring pulmonary congestion and predicting admissions for decompensation. J Clin Monit Comput 2015 Jun;29(3):341-349. [doi: 10.1007/s10877-014-9610-6] [Medline: 25193676]
20. Shochat MK, Shotan A, Blondheim DS, Kazatsker M, Dahan I, Asif A, et al. Non-Invasive Lung IMPEDANCE-Guided Preemptive Treatment in Chronic Heart Failure Patients: A Randomized Controlled Trial (IMPEDANCE-HF Trial). J Card Fail 2016 Sep;22(9):713-722. [doi: 10.1016/j.cardfail.2016.03.015] [Medline: 27058408]
21. Montalibet A, Arkouche W, Bogónez Franco P, Bonnet S, Clarion A, Delhomme G, et al. Design and development of an impedimetric-based system for the remote monitoring of home-based dialysis patients. Stud Health Technol Inform 2016;224:78-83. [Medline: 27225557]
22. Bonnet S, Bourgerette A, Gharbi S, Rubeck C, Arkouche W, Massot B, et al. Wearable impedance monitoring system for dialysis patients. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:5196-5199. [doi: 10.1109/EMBC.2016.7591898] [Medline: 28269435]
23. Montalibet A, Arkouche W, Bogonez Franco P, Bonnet S, Clarion A, Delhomme G, et al. Localised impedance monitoring device for the remote clinical assessment of home-based dialysis patients. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:4427-4430. [doi: 10.1109/EMBC.2016.7591709] [Medline: 28269260]
24. Mabote T, Wong K, Cleland JGF. The utility of novel non-invasive technologies for remote hemodynamic monitoring in chronic heart failure. Expert Rev Cardiovasc Ther 2014 Aug;12(8):923-928. [doi: 10.1586/14779072.2014.935339] [Medline: 25026973]
25. Malfatto G, Corticelli A, Villani A, Giglio A, Della Rosa F, Branzi G, et al. Transthoracic bioimpedance and brain natriuretic peptide assessment for prognostic stratification of outpatients with chronic systolic heart failure. Clin Cardiol 2013 Feb;36(2):103-109 [FREE Full text] [doi: 10.1002/clc.22086] [Medline: 23377871]
26. Beckmann L, Cordes A, Saygili E, Schmeink A, Schauerte P, Walter M, et al. Monitoring of body fluid in patients with chronic heart failure using Bioimpedance - Spectroscopy. In: Dössel O, Schlegel WC, editors. World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany. IFMBE Proceedings, vol 25/7. Berlin, Heidelberg: Springer; 2009:532-535.
27. Dovancescu S, Saczynski JS, Darling CE, Riistama J, Sert Kuniyoshi F, Meyer T, et al. Detecting Heart Failure Decompensation by Measuring Transthoracic Bioimpedance in the Outpatient Setting: Rationale and Design of the SENTINEL-HF Study. JMIR Res Protoc 2015 Oct 09;4(4):e121 [FREE Full text] [doi: 10.2196/resprot.4899] [Medline: 26453479]
28. Darling CE, Dovancescu S, Saczynski JS, Riistama J, Sert Kuniyoshi F, Rock J, et al. Bioimpedance-Based Heart Failure Deterioration Prediction Using a Prototype Fluid Accumulation Vest-Mobile Phone Dyad: An Observational Study. JMIR Cardio 2017 Mar 13;1(1):e1 [FREE Full text] [doi: 10.2196/cardio.6057] [Medline: 31758769]
29. Cuba-Gyllensten I, Gastelurrutia P, Riistama J, Aarts R, Nuñez J, Lupon J, et al. A novel wearable vest for tracking pulmonary congestion in acutely decompensated heart failure. Int J Cardiol 2014 Nov 15;177(1):199-201. [doi: 10.1016/j.ijcard.2014.09.041] [Medline: 25499378]
30. Gastelurrutia P, Cuba-Gyllensten I, Lupon J, Zamora E, Llibre C, Caballero Á, et al. Wearable vest for pulmonary congestion tracking and prognosis in heart failure: A pilot study. Int J Cardiol 2016 Jul 15;215:77-79. [doi: 10.1016/j.ijcard.2016.04.024] [Medline: 27111163]
31. Seulki L, Squillace G, Smeets C, Vandecasteele M, Grieten L, de Francisco R, et al. Congestive heart failure patient monitoring using wearable Bio-impedance sensor technology. Annu Int Conf IEEE Eng Med Biol Soc 2015 Aug;2015:438-441. [doi: 10.1109/EMBC.2015.7318393] [Medline: 26736293]

32. Smeets CJP, Lee S, Groenendaal W, Squillace G, Vranken J, De Cannière H, et al. The Added Value of In-Hospital Tracking of the Efficacy of Decongestion Therapy and Prognostic Value of a Wearable Thoracic Impedance Sensor in Acutely Decompensated Heart Failure With Volume Overload: Prospective Cohort Study. JMIR Cardio 2020 Mar 18;4(1):e12141 [FREE Full text] [doi: 10.2196/12141] [Medline: 32186520]
33. Aamodt IT, Lycholip E, Celutkiene J, von Lueder T, Atar D, Falk RS, et al. Self-Care Monitoring of Heart Failure Symptoms and Lung Impedance at Home Following Hospital Discharge: Longitudinal Study. J Med Internet Res 2020 Jan 07;22(1):e15445 [FREE Full text] [doi: 10.2196/15445] [Medline: 31909717]
34. Michard F. Lung water assessment: from gravimetry to wearables. J Clin Monit Comput 2019 Feb;33(1):1-4. [doi: 10.1007/s10877-018-0154-z] [Medline: 29752666]

Abbreviations

BIA: bioelectrical impedance analysis
CO: cardiac output
ECG: electrocardiography
HF: heart failure
HR: hazard ratio
ICG: impedance cardiography
ITI: internal thoracic impedance
LI: lung impedance
LVEF: left ventricular ejection fraction
NNT: number needed to treat
NYHA: New York Heart Association
STEMI: ST-elevated myocardial infarction
SV: stroke volume

Edited by L Buis; submitted 21.11.20; peer-reviewed by R Piotrowicz, G Cybulski, C Smeets; comments to author 22.01.21; revised version received 07.02.21; accepted 31.03.21; published 05.05.21.

Please cite as:
Krzesinski P, Sobotnicki A, Gacek A, Siebert J, Walczak A, Murawski P, Gielerak G
Noninvasive Bioimpedance Methods From the Viewpoint of Remote Monitoring in Heart Failure
JMIR Mhealth Uhealth 2021;9(5):e25937
URL: https://mhealth.jmir.org/2021/5/e25937
doi:10.2196/25937
PMID:33949964

Review

The Use of Head-Worn Displays for Vital Sign Monitoring in Critical and Acute Care: Systematic Review

Francine Elrose[1], BA(Hons); Andrew Hill[1,2,3], PhD; David Liu[4], MBBS, PhD; Isaac Salisbury[1], BPsySc(Hons); Thien LeCong[5,6], MBBS; Robert G Loeb[1,7], MD; Penelope Sanderson[1,4,8], PhD

[1]School of Psychology, The University of Queensland, St Lucia, QLD, Australia

[2]Minerals Industry Safety and Health Centre, Sustainable Minerals Institute, The University of Queensland, St Lucia, QLD, Australia

[3]Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, QLD, Australia

[4]School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, QLD, Australia

[5]South Australian Ambulance Service (MedSTAR), Adelaide, SA, Australia

[6]The Lyell McEwin Hospital, Elizabeth Vale, SA, Australia

[7]College of Medicine, University of Florida, Gainesville, FL, United States

[8]School of Clinical Medicine, The University of Queensland, Herston, QLD, Australia

Corresponding Author:
Penelope Sanderson, PhD
School of Psychology
The University of Queensland
McElwain Building
St Lucia, QLD
Australia
Phone: 61 407288695
Email: p.sanderson@uq.edu.au

Abstract

Background: Continuous monitoring of patient vital signs may improve patient outcomes. Head-worn displays (HWDs) can provide hands-free access to continuous vital sign information of patients in critical and acute care contexts and thus may reduce instances of unrecognized patient deterioration.

Objective: The purpose of the study is to conduct a systematic review of the literature to evaluate clinical, surrogate, and process outcomes when clinicians use HWDs for continuous patient vital sign monitoring.

Methods: The review was registered with PROSPERO (CRD42019119875) and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines. A literature search was conducted for articles published between January 1995 and June 2020 using the following databases: PubMed, Embase, CINAHL, PsycINFO, and Web of Science. Overall, 2 reviewers independently screened titles and abstracts and then assessed the full text of the articles. Original research articles that evaluated the clinical, surrogate, or process outcomes of head-mounted displays for continuous vital sign monitoring in critical care or acute care contexts were included.

Results: Of the 214 records obtained, 15 (7%) articles met the predefined criteria and were included in this review. Of the 15 studies, 7 (47%) took place in a clinical context, whereas the remainder took place in a simulation environment. In 100% (7/7) of the studies that evaluated gaze behavior, changes were found in gaze direction with HWDs. Change detection improvements were found in 67% (2/3) of the studies evaluating changes in the participants' ability to detect changes in vital signs. Of the 10 studies assessing the ease of use of the HWD, most participants of 7 (70%) studies reported that the HWD was easy to use. In all 6 studies in which participants were asked if they would consider using the HWD in their practice, most participants responded positively, but they often suggested improvements on the HWD hardware or display design. Of the 7 studies conducted in clinical contexts, none reported any clinical outcomes.

Conclusions: Although there is limited and sometimes conflicting evidence about the benefits of HWDs from certain surrogate and process outcomes, evidence for clinical outcomes is lacking. Recommendations are to employ user-centered design when developing HWDs, perform longitudinal studies, and seek clinical outcomes.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019119875; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=119875

(JMIR Mhealth Uhealth 2021;9(5):e27165) doi:10.2196/27165

KEYWORDS

wearable; wearable device; head-mounted display; head-worn display; clinical setting; medical setting; patient monitoring; healthcare

Introduction

Background

Early recognition of patient deterioration can improve patient outcomes [1-3]. Vital sign monitoring helps clinicians track physiological parameters that can provide prodromal warning signs of critical illness [4]. In acute and critical care contexts, vital sign monitoring may reduce preventable in-hospital deaths by improving the early detection of clinical deterioration and allowing for timely intervention to reverse physiological decline [5-7].

There are two approaches to in-hospital vital sign monitoring: intermittent and continuous. Intermittent monitoring involves the measurement and recording of vital signs at regular time intervals (eg, 30 minutes). Some research suggests that some patient deterioration may not be detected because of the gap between observations [8]. Continuous monitoring, recognized as a more proactive approach that is typically used for more at-risk patients, involves the continuous capture of vital sign information or information about significant changes, which is then transmitted to a display device. Although the utility of continuous monitoring for low-intensity patients is still under debate, a recent systematic review focusing on general hospital wards found that continuous vital sign monitoring showed clinical benefits over intermittent monitoring, including reduced critical care use and reduced length of hospital stay, as well as an overall reduction in the cost of patient care [9].

One of the impediments to implement continuous vital sign monitoring is that it can increase the burden on health care workers if algorithms are not used to limit nonactionable or nuisance alarms [10]. The fact that current continuous vital sign monitoring systems have fixed physical locations—at the bedside or at a central station—might also increase the cognitive and physical workload of health care workers when they need to move to the monitor location to see patients' status. Even when monitoring information is displayed nearby, it may be at a location that is awkward to see, such as located behind the anesthesiologist in the operating room. A solution may be the use of wearable devices to provide clinicians with continuous access to patient information, regardless of their location.

Head-worn displays (HWDs) are a type of wearable device that projects information in front of one eye (monocular) or both eyes (binocular) over a background that is either transparent or opaque. HWDs have been trialed in health care contexts for a variety of purposes, such as visual instruction and augmented reality during surgery, videoconferencing between physicians and consultants, and image and video recording for educational purposes [11].

HWDs offer several potential benefits in health care contexts. They allow clinicians to maintain sterility while accessing task-relevant information, which is beneficial in the clinical context [12]. They provide patient information independent of head orientation, which may be particularly advantageous when used for continuous patient monitoring. They present information that is *always there* while clinicians move around the environment and complete other tasks.

Objectives

The purpose of this review is to examine the evidence for the effectiveness of HWDs for continuous patient vital sign monitoring in hospital or patient transport environments. Recent reviews have surveyed the broad range of uses of HWDs in surgery [11,13] or in a wide range of care situations [14]. Other reviews have focused specifically on Google Glass and its uses in surgical environments [15] or nonsurgical environments [16]. This review is not restricted to any specific HWD device. It also offers a more probing and detailed assessment of one specific class of use—the effect of HWD-based vital sign monitoring on outcomes—and it assesses the quality and risk of bias in the papers reviewed. Three types of outcomes are of interest: (1) clinical outcomes (measurable changes in mortality, morbidity, or patient complications), (2) surrogate measures (markers that may correlate with clinical outcomes but do not have a guaranteed relationship, such as detection of deterioration, or reduction in time to detect vital sign changes), and (3) process measures (procedural aspects related to the clinical process, such as changes in the pattern of gaze changes, or improvements in information sharing).

Methods

Overview

This systematic review was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines. The review was registered with the International Prospective Register of Systematic Reviews (CRD42019119875) before the search process started.

Article Retrieval

We developed the search terms with the assistance of a professional librarian at The University of Queensland. The search strategy used keywords related to the concepts of *head-worn displays*, *vital signs*, and *patients*. The full search terms are included in Table S1 in Multimedia Appendix 1. The searches were run in the following databases in June 2019 and again in June 2020: PubMed, Embase, CINAHL, PsycINFO, and Web of Science. Search strategies were adapted for each database to allow for differences in the required search techniques. The search strategies are listed in Table S2 in Multimedia Appendix 1. Once the articles were selected, their reference lists were screened to identify eligible articles not located using the search strategy; these articles were then screened based on the same inclusion and exclusion criteria.

Selection Criteria

The inclusion criteria were as follows: (1) peer-reviewed qualitative, quantitative, or mixed methods studies (excluding gray literature, editorials, systematic or other reviews, and meta-analyses); (2) real or simulated critical care or acute care clinical contexts; (3) fully trained clinician or clinical trainee participants; (4) HWDs used for vital sign monitoring (excluding images or videos without vital sign information); (5) studies with or without a comparator; and (6) predefined clinical, surrogate, or process outcomes. The list of outcomes was developed prospectively in collaboration with 2 practicing clinicians to focus on outcomes relevant to the clinical context and is provided in Table S3 in Multimedia Appendix 1. The full inclusion and exclusion criteria are also included in Table S4 in Multimedia Appendix 1.

Data Extraction and Analysis

A standardized form was developed for the structured collection of data from each article, including publication details, study methodology, HWD details, and study outcomes; full details of the data collected are presented in Table S5 in Multimedia Appendix 1. For the study outcomes, classifications were made based on the descriptions of the outcomes by authors of the retrieved articles. Overall, 2 authors (FE and PS) independently extracted data from all the articles. Discrepancies were resolved through discussion with a third author (AH) when necessary. The data extracted from the articles were analyzed both quantitatively and qualitatively.

Risk of Bias Assessment

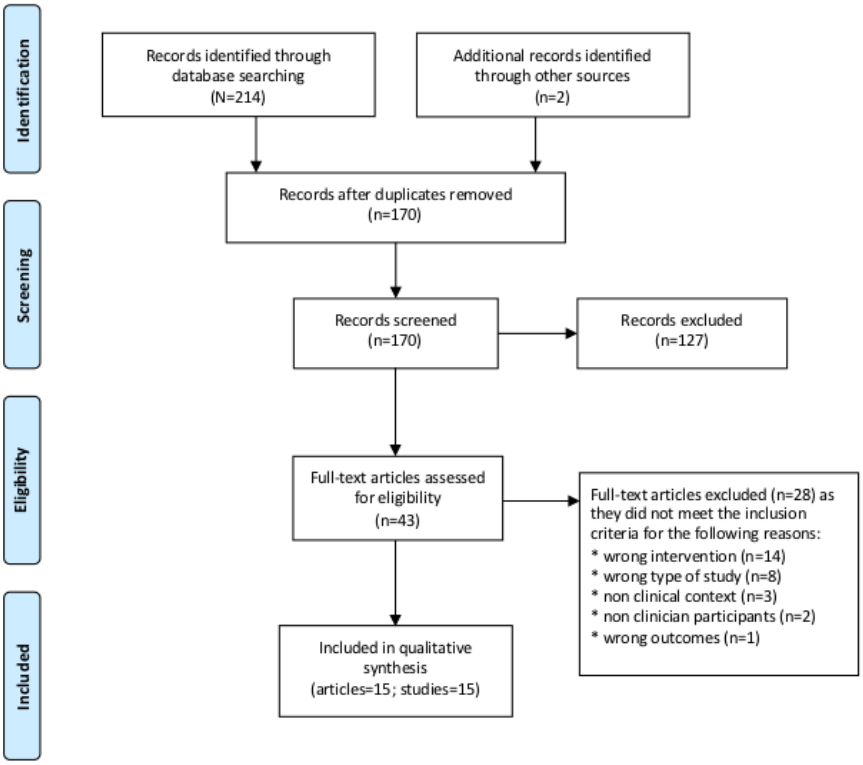
The risk of bias assessment considers the extent to which the design of a study and methods used are likely to have prevented bias [17]. Two authors (FE and PS) independently assessed the risk of bias of each study using the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields [18]. According to Kmet et al [18], each study is rated according to whether it meets specific criteria with *yes* (2 points), *partial* (1 point), or *no* (0 points), with a *not applicable* ("N/A") option available for selected criteria. A summary score is then calculated across all relevant items as a percentage of the possible total score. A higher percentage represents an assessment of higher-quality research and lower risk of bias. Where a study uses both quantitative and qualitative methodologies, it is assessed against both criteria, and 2 summary scores are calculated.

Results

Literature Search

A total of 214 citations were retrieved through a structured search of the 5 databases. Two additional records were identified through other sources. After removing duplicates, 170 articles were screened. In total, 2 authors (FE and IS) independently screened the titles and abstracts of 170 records against the inclusion and exclusion criteria, and 43 articles met all criteria. The reviewers were in agreement for 97.6% (166/170) records, and discrepancies were resolved through discussion with a third author (PS). The initial 2 authors (FE and IS) independently reviewed the full text of these articles against the inclusion and exclusion criteria, and 28 articles were excluded. The reviewers were in agreement for 100% (43/43) of the articles. A total of 15 articles met the predefined criteria for inclusion. Note that one article described 2 separate studies [19], and 2 articles described the same study but with a different analytic focus [20,21]; therefore, 15 studies were included in the review. The study flowchart and the reasons for exclusion at each stage are documented in the PRISMA study flowchart in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) study flowchart.



Description of Study Characteristics

Table 1 summarizes the 15 included studies. Most studies used a crossover design (10/15, 67%) [19-28], whereas the remainder used a case series design (3/15, 20%) [29-31] or a case study design (2/15, 13%) [32,33]. Approximately half of the studies (8/15, 53%) occurred in a simulated clinical context [19,22-27].

Approximately half of the studies (7/15, 47%) used a volunteer or convenience participant sample [19-21,24,26,29,30], whereas the remainder did not specify how their sample was recruited [22,23,25,27,28,31-33]. The sample size ranged from 2 to 40 participants (median 11.00; mean 12.60, SD 11.35); one study did not report the number of participants (1/15, 7%) [25]. Most of the studies (9/15, 60%) recruited anesthesiologists as participants [19-21,25,26,28-31], and a few studies (3/15, 20%) recruited surgeons [22,24,32]. Most of the studies reported quantitative information about the participants' medical experience (10/15, 67%) [19,21,23,24,26-28,30,33], and a few studies reported the ages of the participants (3/15, 20%) [24,27,28].

Table 1. Summary of the studies included.

Study	Study design	Participants (n); context	Comparison condition	Type of head-worn display	Format of data;source[a]
Beuchat et al, 2005 [22]	Randomized controlled trial with crossover design	Cardiovascular surgeons (n=4); simulated operating room (open heart surgery)	Standard monitoring	Binocular; opacity unspecified; Sony Glasstron	Waveforms and numbers; mirror
Block et al, 1995 [29]	Case series	Anesthesiologists (n=11); operating room (anesthesia)	None	Monocular optical see-through; Reflection Technology Private Eye	Numbers; redesign
Drake-Brockman et al, 2016 [30]	Case series	Anesthesiologists—pediatric (n=40); operating room (anesthesia)	None	Monocular optical see-through; Google Glass	Numbers; redesign
Iqbal et al, 2016 [23]	Crossover trial with fixed order of presentation	Urologists (n=37); simulated operating room (prostatectomy)	Standard monitoring	Monocular optical see-through; Google Glass	Waveforms and numbers; not stated
Liebert et al, 2016 [24]	Randomized controlled trial with crossover design	Surgical residents (n=14); simulated operating room (thoracostomy, bronchoscopy)	Standard monitoring	Monocular optical see-through; Google Glass	Waveforms and numbers; mirror
Liu et al, 2009 [20] and Liu et al, 2010 [21][b]	Randomized controlled trial with crossover design	Anesthesiologists (n=6); operating room (anesthesia)	Standard monitoring	Monocular optical see-through; Microvision Nomad	Waveforms and numbers; limited replica
Liu et al, 2009 [19]: experiment 1	Crossover design with Latin square assignment to order	Anesthesiologists (n=12); simulated operating room (anesthesia)	Standard monitoring	Monocular optical see-through; Microvision Nomad	Waveforms and numbers; redesign
Liu et al, 2009 [19]: experiment 2	Crossover design with alternating allocation to conditions	Anesthesiologists (n=12); simulated operating room (anesthesia)	Standard monitoring	Monocular optical see-through; Microvision Nomad	Waveforms and numbers; redesign
Ormerod et al, 2002 [25]	Crossover trial with alternative allocation to condition	Anesthesiologists (sample not stated); simulated operating room (anesthesia)	Standard monitoring	Monocular optical see-through; Microvision Nomad	Waveforms and numbers; not stated
Sanderson et al, 2008 [26]	Crossover trial with Latin square assignment to order	Anesthesiologists (n=16); simulated operating room (anesthesia)	Standard monitoring	Monocular optical see-through; Microvision Nomad	Numbers; redesign
Schaer et al, 2015 [27]	Randomized controlled trial with crossover design	Medical residents (n=7); simulated surgical setting (cardiac surgery)	Standard monitoring	Monocular optical see-through; Google Glass	Waveforms and numbers; redesign
Schlosser et al, 2019 [28]	Randomized controlled trial with crossover design	Anesthesiologists—supervising (n=6); operating suite (multiple patient anesthesia)	Standard monitoring	Monocular opaque; Vuzix M300	Waveforms and numbers; redesign
Via et al, 2002 [31]	Case series	Anesthesiologists (n=12); operating room (anesthesia)	None	Binocular optical see-through; Kaiser Electro-Optics	Waveforms and numbers; mirror
Vorraber et al, 2014 [32]	Case study	Surgeons (n=2); operating room (percutaneous transluminal angioplasty)	None	Monocular optical see-through; Google Glass	Waveforms and numbers; mirror
Yoshida et al, 2014 [33]	Case study	Urologists (n=2); operating room (transurethral resection of the prostate)	None	Binocular opaque; Sony HMZ-T2	Waveforms and numbers; redesign

[a]Source refers to whether the *head-worn display* showed vital sign information in the same format as the standard monitor (mirror), a similar format with some vital sign information removed (limited replica), or in a new format (redesign).

[b]Liu et al's clinical study outcomes were reported over 2 papers; treatment here integrates findings from both studies [20,21].

Description of HWD Type and Interface

Table 1 summarizes the types of HWD and how the data were displayed. Most of the studies (12/15, 80%) used monocular HWDs [19-21,23-30,32], and the remainder used binocular HWDs (3/15, 20%) [22,31,33]. Of the studies using monocular HWDs, 8% (1/12) used an opaque HWD [28], whereas the remaining studies used optical see-through HWDs (11/12, 92%). Of the 3 studies using binocular HWDs, the HWD in one study (1/3, 33%) was opaque [33], creating an immersive virtual reality experience; in another study, it was optical see-through

(1/3, 33%) [31]; and the third study did not specify whether the background was opaque or transparent (1/3, 33%) [22].

Of 15 studies, 12 (75%) displayed a combination of waveforms and numbers on the HWD [19-25,27,28,31-33] and 3 (20%) presented only numbers [26,29,30]. Only 27% (4/15) of the studies presented identical data in the same format on the HWD as on the standard patient monitor [22,24,31,32]. Of 15 studies, 1 (7%) presented on the HWD a subset of the data available on the standard patient monitor [20,21]. Approximately half (8/15, 53%) of the studies presented identical data on the HWD as on the standard monitor but in a redesigned format [19,26-30,33]. Of 15 studies, 2 (13%) studies did not state whether the data were in the same format as the patient monitor or in a redesigned format [23,25], and only 1 (7%) study conducted a formal requirements analysis before implementing and testing the HWD [32].

Risk of Bias of Included Studies

For the 13 studies with quantitative measures, quality ratings [18] ranged from 25% to 92%, with a median rating of 75%. For the 3 studies with qualitative measures, quality ratings ranged from 30% to 75%, with a median rating of 39%. Detailed quality assessments are included in Table S6 in Multimedia Appendix 1.

Description of Findings for Clinical Outcomes

Table 2 shows the outcomes extracted from each study. A table with the same data in an extended format is shown in Table S7 in Multimedia Appendix 1. We examined papers for any reports of 3 patient-related clinical outcomes: mortality, morbidity, and rate of complications. Although 7 studies were performed with human patients in clinical contexts rather than simulated patients [21,28-33], none of these studies provided any data relating to the clinical outcomes mentioned earlier.

Table 2. Summary of the clinical, surrogate, and process outcomes considered in each study.

Study	Clinical outcomes	Surrogate outcome categories[a]	Process outcome categories[a]	Quality rating (%)	
				Quantitative measure	Qualitative measure
Beuchat et al, 2005 [22]	N/A[b]	+ Decreased missed vital sign changes + Reduced time to detect vital sign changes	+ Changed patterns of gaze behavior + Increased time focused on patient	58	—[c]
Block et al, 1995 [29]	N/M[d]	N/M	+ Clinician opinions	25	—
Drake-Brockman et al, 2016 [30]	N/M	N/M	+ Clinician opinions	83	—
Iqbal et al, 2016 [23]	N/A	+ Reduced time to detect vital sign changes	= Changed time to do other tasks + Clinician opinions	63	—
Liebert et al, 2016 [24]	N/A	= Earlier identification of deterioration	+ Changed patterns of gaze behavior + Increased time focused on patient	92	—
Liu et al, 2009 [20] and Liu et al, 2010 [21][e]	N/M	N/M	+= Changed patterns of gaze behavior + Increased time focused on patient = Clinician opinions	88	—
Liu et al, 2009 [19]: experiment 1	N/A	= Increased unexpected events detected = Reduced time to detect unexpected events	+ Changed patterns of gaze behavior + Increased time focused on patient = Clinician opinions	75	—
Liu et al, 2009 [19]: experiment 2	N/A	+= Increased unexpected events detected += Reduced time to detect unexpected events	+ Changed patterns of gaze behavior + Increased time focused on patient + Clinician opinions	75	—
Ormerod et al, 2002 [25]	N/A	N/M	+ Changed patterns of gaze behavior + Increased time focused on patient + Changed time to do other tasks + Clinician opinions	29	—
Sanderson et al, 2008 [26]	N/A	= Increased unexpected events detected = Reduced time to detect unexpected events	+ Clinician opinions	83	—
Schaer et al, 2015 [27]	N/A	N/M	+ Clinician opinions	71	—
Schlosser et al, 2019 [28]	N/M	+ Increased alarms detected = Reduced time to detect alarms	+− Clinician opinions	92	75
Via et al, 2002 [31]	N/M	N/M	+ Clinician opinions	59	—
Vorraber et al, 2014 [32]	N/M	N/M	+ Changed patterns of gaze behavior + Clinician opinions	—	30
Yoshida et al, 2014 [33]	N/M	N/M	+ Clinician opinions	—	39

[a]"+" represents the positive effect of head-worn display; "−" represents the negative effect of head-worn display; "=" represents no difference between head-worn display and other conditions. If a study used more than one measure for an outcome, multiple symbols are shown.

[b]N/A: not applicable.

[c]Not relevant to this study.

[d]N/M: not measured.

[e]Liu et al's clinical study outcomes were reported over 2 papers; treatment here integrates findings from both studies [20,21].

Description of Findings for Surrogate Outcomes

We extracted data for 9 predetermined surrogate outcomes that were subsequently divided into 4 categories: (1) vital sign changes, (2) alarm detection, (3) unexpected event detection, and (4) situation awareness.

Vital Sign Changes

Of 15 studies, 3 (20%) examined the participants' ability to detect vital sign changes in simulated surgical settings. Overall, 2 of the studies found that the average time to detect abnormal vital signs was significantly faster with an HWD than with standard monitoring. In the first study, Beuchat et al [22] found that participants using an HWD responded almost twice as fast as participants using conventional monitoring, although the overall duration of the surgical intervention was the same for HWD and conventional monitoring. Beuchat et al [22] also found that abnormal vital signs were always detected by participants with an HWD but not by participants using conventional monitoring.

In the second study, Iqbal et al [23] reported that 84% of participants using the HWD responded faster to abnormal vital signs than participants in the conventional monitoring group; however, measures of technical performance from the simulation, including a measure of simulated blood loss, were similar for participants across groups. In the third study, Liebert et al [24] found only a nonsignificant trend for participants using the HWD to recognize abnormal vital signs faster than participants using a standard monitor. Overall, there is some evidence that HWDs can improve participants' detection of vital sign changes.

Alarms

One study examined whether the use of an HWD increased the number of auditory alarms detected by clinicians or reduced the time taken for clinicians to detect alarms. Schlosser et al [28] delivered alarms in visual and auditory formats via the HWD and found that anesthesiologists supervising work in 6 operating rooms noticed a higher percentage of alarms when using the HWD than when relying on central monitoring or on monitoring within each operating room (67% compared with 7%). There was no difference between the conditions in the median time taken to detect alarms.

Unexpected Events

Of 15 studies, 3 (20%) examined the effect of HWDs on the detection of unexpected events and the time taken to respond to these events. For unexpected events occurring on the HWD, such as hypertension or gas embolism, 2 simulator studies found that participants using the HWD did not detect more unexpected events [26] and did not detect them faster than participants using standard monitoring [19,26]. In a third simulator study, Liu et al [19] (experiment 2), anesthesiologists worked under conditions where they were either operationally constrained (by being involved in an intubation task) or physically constrained (by requiring 2 hands to operate a surgical tool). The HWD increased the detection rate and speed of detection in some clinical scenarios (tachypnea or hypertension and hypotension events) but not others (hypoventilation), compared with standard monitoring.

Liu et al [19] also examined unexpected events occurring outside the HWD, such as the simulated patient opening their eyes while anesthetized or a medical student fainting. Participants using the HWD did not detect events faster than the participants using standard monitoring. Overall, the HWD does not appear to improve participants' ability to notice unexpected events unless participants are constrained from noticing them on a standard monitor.

Situation Awareness

Our intended criterion for including situation awareness outcomes was that a validated measure should be used; none of the studies met this criterion. However, one study described nonvalidated measures of situation awareness, specifically self-reported awareness, and are reported here for completeness. Schlosser et al [28] surveyed clinicians after they used an HWD and found that they reported that the HWD helped them to comprehend the environment and make assessments more easily and that information gained from the HWD affected future actions.

Description of Findings for Process Outcomes

We extracted data for 5 predetermined process outcomes that were divided into 4 categories: (1) gaze behavior, (2) time for other tasks, (3) information sharing, and (4) clinicians' opinions of HWDs.

Gaze Behavior

Of 15 studies, 7 (47%) examined the gaze behavior of clinicians using an HWD. Of 7 studies, 5 found that clinicians using an HWD spent more time looking at the patient or the procedural field and less time looking at the monitor than when the HWD was not used [19,21,24,25,32] and 2 found that clinicians using an HWD showed a decrease in head movements or shifts in attention, compared with those using standard monitoring [22,25]. A further analysis of the data in a study by Liu et al [21] indicated that when using an HWD, clinicians spent more time looking toward the patient and less time looking toward an anesthesia machine than when they used standard monitoring for some clinical scenarios (eg, anesthesia crisis management) but not for other scenarios (anesthesia confirmation of laryngeal mask airway placement) [20].

Time Spent on Tasks

Overall, 2 studies examined the effect of using an HWD on the time spent on tasks. In one study, Ormerod et al [25] found that clinicians using an HWD spent less time completing anesthesiology tasks than clinicians using standard monitoring, with the implication that the HWD removed the need for clinicians to interrupt tasks to view a monitor. In another study, Iqbal et al [23] found that clinicians using an HWD performed a surgical task at a similar speed to clinicians using a standard monitor.

Information Sharing

None of the studies considered the effect of HWDs on how clinicians share information.

Clinicians' Opinions

Of 15 studies, 14 (93%) examined some aspects of clinicians' opinions of HWDs for vital sign monitoring [19,21,23-33]. Overall, clinicians had positive opinions about the use of HWDs. In 10 studies, clinicians were asked about the ease of reading or interpretation; in 7 (70%) of those studies, most clinicians reported that the HWD was easy to read or interpret [19,24-28,30]. However, one study found no difference in ease of monitoring between clinicians using HWDs and clinicians using standard monitoring [21], and one study found that approximately half of the clinicians preferred an HWD whereas half preferred using the standard monitor [19].

In 6 studies, clinicians were asked whether wearing the HWD was comfortable. In 5 of these studies, most clinicians reported that the HWD was comfortable to wear or did not report any discomfort [23,29,30,32,33]. However, in one study, half of the clinicians reported that the HWD was too big or too heavy, and they experienced discomfort or pain from wearing it [28].

Finally, in all 6 studies where clinicians were asked whether they would use the device again, most clinicians said yes, although sometimes noting improvements needed to the device or to the information presented [23,24,28-31].

Discussion

Principal Findings

The purpose of this review was to evaluate the impact of HWDs displaying continuous vital sign monitoring on clinical, surrogate, and process outcomes in critical and acute care contexts. Our systematic review of the literature shows that HWDs have been evaluated for continuous vital sign monitoring in 15 studies, including 7 conducted with patients in clinical environments. Clearly, HWDs can be technically implemented for vital sign monitoring, but the evidence for any overall benefit is mixed. None of the 7 clinical studies measured clinical outcomes. Across all 15 studies, there was only limited evidence that HWDs displaying patient vital signs improve surrogate outcomes or process outcomes.

The strongest and most consistent evidence for the benefit of HWDs relates to gaze behavior. Several studies have shown that wearing an HWD tends to increase the time that clinicians spend looking toward the patient relative to the patient monitor or anesthesia machine [19-22,24,25,32]. However, as Liu et al [19] pointed out, when the HWD image overlays the view of a patient, a clinician may be looking toward a patient but directing their attention to the HWD image rather than to the patient. Therefore, although the evidence for changes in gaze behavior toward the patient is consistent, it does not necessarily imply greater attention to the patient.

There is also consistent evidence that clinicians using an HWD may take less time to detect abnormal vital signs and may also detect more vital sign changes than clinicians using standard monitoring. However, this evidence comes from a limited number of studies—2 with significant results [22,23] and one with a nonsignificant trend [24]—and is therefore not conclusive.

Evidence for whether HWDs affect clinicians' detection of unexpected events signaled on an HWD is mixed. In 2 studies, there was no difference between HWDs and standard monitoring for detecting unexpected events [19,26]. Liu et al [19] reported that in operationally constrained conditions, HWDs may increase the detection rate and speed of detection of unexpected events compared with standard monitoring. Specifically, when vital signs indicating an unexpected event were presented as numbers, the HWD increased the detection of unexpected events. However, when the relevant vital sign was presented as a waveform, the HWD worsened detection. This suggests that when vital sign information on the HWD is in the forward field of view, certain formats may actually be less detectible compared with the same information presented on a standard monitor. Liu et al [19] found no difference between HWDs and standard monitoring in how often clinicians noticed unexpected events occurring in the simulated operating room rather than on the HWD.

For other outcomes, the evidence is mixed. HWDs are sometimes associated with improvements in alarm detection and the time required to complete tasks, but sometimes not. This makes it difficult to provide definitive conclusions regarding the effects that HWDs have on these outcomes. Further research is needed to determine the factors that might moderate the effect of HWDs on these outcomes. For situation awareness, the fact that no study has collected objective measures of the 3 levels of situation awareness by Endsley [34,35]—how HWDs affect clinicians' perception and comprehension of the patient's current state and their projection of the patient's future state—means we do not yet know the impact of HWDs on situation awareness. Subjective and observational evidence suggests that HWDs may improve clinicians' awareness of the ongoing situation [28], but further research is needed to verify this claim.

Future Research

For more definitive conclusions to be drawn about the impact of HWDs on vital sign monitoring, the focus and quality of the research need to be improved. The quality assessment ratings ranged from 25% to 92%, indicating that there is room for improvement in study design and methods. Given that there is no strong evidence that HWDs worsened performance on any outcome measures reported, it is worth investing in higher-fidelity studies to improve the evidence base.

The inconsistent results may be partly explained by the use of small-scale, short-term studies with a focus on surrogate and process outcomes. Longitudinal studies may allow clinicians to adjust to the novelty of using an HWD and may reveal whether the benefits of HWDs emerge over time. Large-scale randomized control trials have not yet emerged, but clinically based studies focusing on clinically relevant outcomes will help clinical leaders decide whether HWDs should be implemented on a large-scale basis.

In the literature reviewed, there was little apparent attention to user-centered design principles [36] when researchers designed and tested the HWD display format and content for critical and acute care environments. A user-centered design approach would help developers to focus on the outcomes that the HWD might improve. The implementation of user-centered design principles would involve determining the requirements that the system (ie, the HWD) must meet to achieve the specified outcomes [37], designing the system to meet those requirements [38], and then conducting empirical tests with users to determine if the requirements have indeed been met and outcomes, achieved. The process is repeated until the requirements are met. Given that users' opinions about a device or technology can sometimes differ markedly from their performance when using that device or technology [39-41] and that it is the performance that will address the outcomes, it is essential to perform robust empirical tests.

No study in this literature review reported a systematic analysis of user needs. A requirements analysis was reported in only one of the 15 studies (Vorraber et al [32]) before designing and testing the HWD system. The largely positive feedback from clinicians reported in the studies reviewed may indicate that they discern uses and usefulness of HWDs that have not yet materialized in practice or they may be overlooking requirements that, without a prospective requirements analysis, would only emerge if HWDs were fully integrated into practice.

An examination of the HWD interfaces used in the literature also supports the assertion that a user-centered design approach may be worthwhile in future research. Of the 13 studies that reported the layout of the HWD interface, 4 presented the data in the same format on the HWD as on the standard patient monitor [22,24,31,32] and one provided a limited replica [20,21]. The remaining 8 studies redesigned the format of the information [19,26-30,33]. Interestingly, the study by Vorraber et al [32], which included a requirements analysis, used the same format as the standard patient monitor. However, this does not indicate that the same format is appropriate for every clinical context. In studies that used a redesigned format, clinicians tended to prefer an HWD display layout in the same format as the standard monitor, but this was not always the case. In one study, most clinicians reported that they would use the HWD again but that they wanted more information on the display [30]. In another study, 82% of clinicians said they would use the HWD again if it displayed waveforms and numbers in the same format as the standard monitor [29]. However, in a further study, clinicians reported that they found information on the HWD redundant with the information on the standard monitor [26]. It may be that each of these studies represents a different use case. If so, performing a requirements analysis motivated by desired outcomes would guide the design of HWD software and hardware for specific clinical contexts. When coupled with longitudinal evaluation studies, this approach could lead to clearer and more consistent results.

Evaluations of the impact of HWDs would be improved if researchers include tests that allow evidence to accumulate for or against the key outcomes listed in this review. For example, relatively few studies have tested whether HWDs render participants more or less able to detect changes in patient vital signs or to notice unexpected events. If future studies were to deliberately augment evidence for or against key claims, clinical leaders would be able to make more confident decisions about the viability of HWDs.

Strengths and Limitations of the Review

The purpose of this systematic review was to evaluate the benefits or otherwise of HWDs for continuous vital sign monitoring, using a set of predefined outcome categories. The strengths of the review are its scope and the breadth of outcomes considered for each study, revealing the considerable heterogeneity of approaches and findings in the area, and indicating areas for further research.

This review has several limitations. First, the heterogeneous nature of the literature poses a challenge to interpretation. Our search terms may not have captured all relevant publications if different key terms are used across different clinical contexts. However, we tried to address this through an iterative process for developing search terms in collaboration with librarians and clinicians. Second, the quality of evidence ranged considerably across the included studies. Third, there may be evidence in other contexts that HWDs can improve or worsen performance that is relevant to continuous vital sign monitoring in acute and critical care contexts; however, finding such cases was outside the scope of the review. Finally, there are other ways to convey continuously captured vital signs to clinicians. For example, auditory, haptic, or multimodal displays may meet clinicians' requirements for continuous patient monitoring. However, this was beyond the scope of this review.

Conclusions

Certain surrogate and process outcomes suggest that HWDs may assist continuous vital sign monitoring, but to date, there have been no evaluations of whether HWDs improve clinical outcomes. The most consistent evidence across the corpus of studies reviewed is that HWDs can improve clinicians' detection of vital sign changes and reduce the time clinicians spend looking at the patient monitor. However, for other surrogate and process outcomes, the evidence is mixed. A user-centered design approach can produce designs and evaluations that are more focused on the desired outcomes. Further research is required to determine whether, in what contexts, and under what conditions HWDs can reliably support the early recognition of patient deterioration and potentially reduce patient harm.

Acknowledgments

The authors thank Dr Andrew Perry and Mr Stuart Keynes from SAAS MedSTAR for helpful conversations at the outset of this research, Miranda Newell, Librarian, from The University of Queensland Library for help with formulating the searches, and members of the Cognitive Engineering Research Group who gave feedback on an early version of the manuscript. This research was supported by Australian Research Council Discovery Project DP180103702.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material including search terms, search strategy, a full list of outcome questions, study inclusion and exclusion criteria, data extraction categories, the full quality assessment for all included studies, and a summary table of the clinical, surrogate, and process outcomes assessed in each study.
[DOCX File , 51 KB - mhealth_v9i5e27165_app1.docx]

References

1. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004] [Medline: 24342543]
2. Cardoso LT, Grion CM, Matsuo T, Anami EH, Kauss IA, Seko L, et al. Impact of delayed admission to intensive care units on mortality of critically ill patients: a cohort study. Crit Care 2011;15(1):R28 [FREE Full text] [doi: 10.1186/cc9975] [Medline: 21244671]
3. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
4. Ridley S. The recognition and early management of critical illness. Ann R Coll Surg Engl 2005 Sep;87(5):315-322 [FREE Full text] [doi: 10.1308/003588405X60669] [Medline: 16176687]
5. Abenstein JP, Narr BJ. An ounce of prevention may equate to a pound of cure: can early detection and intervention prevent adverse events? Anesthesiology 2010 Feb;112(2):272-273 [FREE Full text] [doi: 10.1097/ALN.0b013e3181ca858d] [Medline: 20098123]
6. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. N Engl J Med 2009 Oct;361(14):1368-1375. [doi: 10.1056/nejmsa0903048]
7. International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. Br J Anaesth 2016 Oct 31;117(5):601-609 [FREE Full text] [doi: 10.1093/bja/aew316] [Medline: 27799174]
8. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth 2006 Jul;97(1):64-68 [FREE Full text] [doi: 10.1093/bja/ael113] [Medline: 16707529]
9. Downey CL, Chapman S, Randell R, Brown JM, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: a systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27. [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
10. Taenzer AH, Pyke JB, McGrath SP, Blike GT. Impact of pulse oximetry surveillance on rescue events and intensive care unit transfers: a before-and-after concurrence study. Anesthesiology 2010 Feb;112(2):282-287 [FREE Full text] [doi: 10.1097/ALN.0b013e3181ca7a9b] [Medline: 20098128]
11. Rahman R, Wood ME, Qian L, Price CL, Johnson AA, Osgood GM. Head-mounted display use in surgery: a systematic review. Surg Innov 2020 Feb 12;27(1):88-100. [doi: 10.1177/1553350619871787] [Medline: 31514682]
12. Iqbal MH, Aydin A, Brunckhorst O, Dasgupta P, Ahmed K. A review of wearable technology in medicine. J R Soc Med 2016 Oct;109(10):372-380 [FREE Full text] [doi: 10.1177/0141076816663560] [Medline: 27729595]
13. Kolodzey L, Grantcharov PD, Rivas H, Schijven MP, Grantcharov TP. Wearable technology in the operating room: a systematic review. BMJ Innov 2016 Dec 21;3(1):55-63. [doi: 10.1136/bmjinnov-2016-000133]
14. Romare C, Skär L. Smart glasses for caring situations in complex care environments: scoping review. JMIR Mhealth Uhealth 2020 Apr 20;8(4):- [FREE Full text] [doi: 10.2196/16055] [Medline: 32310144]
15. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
16. Dougherty B, Badawy SM. Using Google glass in nonsurgical medical settings: systematic review. JMIR Mhealth Uhealth 2017 Oct 19;5(10):e159 [FREE Full text] [doi: 10.2196/mhealth.8671] [Medline: 29051136]
17. Higgins JP, Thomas J. In: Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions, 2nd Edition. New York: John Wiley & Sons; 2019:1-728.
18. Kmet LM, Lee RC, Cook LS, Alberta Heritage Foundation for Medical Research, AHFMR. Health Technology Assessment Unit, University of Calgary. Faculty of Medicine, Calgary Health Region. Standard Quality Assessment Criteria for

Evaluating Primary Research Papers From a Variety of Fields. Edmonton, Alberta: Alberta Heritage Foundation for Medical Research; 2004.
19. Liu D, Jenkins SA, Sanderson PM, Watson MO, Leane T, Kruys A, et al. Monitoring with head-mounted displays: performance and safety in a full-scale simulator and part-task trainer. Anesth Analg 2009;109(4):1135-1146. [doi: 10.1213/ane.0b013e3181b5a200]
20. Liu D, Jenkins SA, Sanderson PM. Clinical implementation of a head-mounted display of patient vital signs. In: Proceedings of International Symposium on Wearable Computers. 2009 Presented at: 2009 International Symposium on Wearable Computers; Sept 4-7, 2009; Linz, Austria p. 47-54. [doi: 10.1109/iswc.2009.36]
21. Liu D, Jenkins SA, Sanderson PM, Fabian P, Russell WJ. Monitoring with head-mounted displays in general anesthesia: a clinical evaluation in the operating room. Anesth Analg 2010 Apr 01;110(4):1032-1038. [doi: 10.1213/ANE.0b013e3181d3e647] [Medline: 20357147]
22. Beuchat A, Taub S, Saby JD, Dierick V, Codeluppi G, Corno AF, et al. Cybertools improve reaction time in open heart surgery. Eur J Cardiothorac Surg 2005 Feb;27(2):266-269. [doi: 10.1016/j.ejcts.2004.10.051] [Medline: 15691680]
23. Iqbal MH, Aydin A, Lowdon A, Ahmed HI, Muir GH, Khan MS, et al. The effectiveness of Google glass as a vital signs monitor in surgery: a simulation study. Int J Surg 2016 Dec;36(Pt A):293-297 [FREE Full text] [doi: 10.1016/j.ijsu.2016.11.013] [Medline: 27833004]
24. Liebert CA, Zayed MA, Aalami O, Tran J, Lau JN. Novel use of Google glass for procedural wireless vital sign monitoring. Surg Innov 2016 Aug;23(4):366-373. [doi: 10.1177/1553350616630142] [Medline: 26848138]
25. Ormerod DF, Ross B, Naluai-Cecchini A. Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. In: Proceedings of the International Symposium on Wearable Computers. 2002 Presented at: International Symposium on Wearable Computers; Oct 10, 2002; Seattle, WA, USA p. 131-132. [doi: 10.1109/iswc.2002.1167229]
26. Sanderson PM, Watson MO, Russell WJ, Jenkins S, Liu D, Green N, et al. Advanced auditory displays and head-mounted displays: advantages and disadvantages for monitoring by the distracted anesthesiologist. Anesth Analg 2008 Jun;106(6):1787-1797. [doi: 10.1213/ane.0b013e31817325cb] [Medline: 18499611]
27. Schaer R, Salamin F, Toro OA, Atzori M, Muller H, Widmer A. Live ECG readings using Google glass in emergency situations. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:315-318. [doi: 10.1109/EMBC.2015.7318363] [Medline: 26736263]
28. Schlosser PD, Grundgeiger T, Sanderson PM, Happel O. An exploratory clinical evaluation of a head-worn display based multiple-patient monitoring application: impact on supervising anesthesiologists' situation awareness. J Clin Monit Comput 2019 Dec 5;33(6):1119-1127. [doi: 10.1007/s10877-019-00265-4] [Medline: 30721389]
29. Block FE, Yablok DO, McDonald JS. Clinical evaluation of the 'head-up' display of anesthesia data. Preliminary communication. Int J Clin Monit Comput 1995 Feb;12(1):21-24. [doi: 10.1007/BF01142442] [Medline: 7782663]
30. Drake-Brockman TF, Datta A, von Ungern-Sternberg BS. Patient monitoring with Google glass: a pilot study of a novel monitoring technology. Paediatr Anaesth 2016 May;26(5):539-546. [doi: 10.1111/pan.12879] [Medline: 26992465]
31. Via D, Kyle R, Geiger P, Mongan P. A head-mounted display of anesthesia monitoring data is of value and would be used by a majority of anesthesia providers. Anesth Analg 2002;95(2 suppl):S132.
32. Vorraber W, Voessner S, Stark G, Neubacher D, DeMello S, Bair A. Medical applications of near-eye display devices: an exploratory study. Int J Surg 2014 Dec;12(12):1266-1272 [FREE Full text] [doi: 10.1016/j.ijsu.2014.09.014] [Medline: 25462704]
33. Yoshida S, Kihara K, Takeshita H, Fujii Y. A head-mounted display-based personal integrated-image monitoring system for transurethral resection of the prostate. Wideochir Inne Tech Maloinwazyjne 2014 Dec;9(4):644-649 [FREE Full text] [doi: 10.5114/wiitm.2014.43040] [Medline: 25562008]
34. Endsley MR. Measurement of situation awareness in dynamic systems. Hum Factors 2016 Nov 23;37(1):65-84. [doi: 10.1518/001872095779049499]
35. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 2016 Nov 23;37(1):32-64. [doi: 10.1518/001872095779049543]
36. Sharp H, Preece J, Rogers Y. Interaction Design: Beyond Human-Computer Interaction, 5th Edition. New York: John Wiley & Sons Inc; 2019:1-656.
37. Grady J. System Requirements Analysis, 2nd Edition. Amsterdam: Elsevier; 2013:1-834.
38. Anton AI. Goal-based requirements analysis. In: Proceedings of the Second International Conference on Requirements Engineering. 1996 Presented at: Second International Conference on Requirements Engineering; April 15-18, 1996; Colorado Springs, CO, USA.
39. Andre AD, Wickens CD. When users want what's not best for them. Ergonomics Design 2016 Jul 21;3(4):10-14. [doi: 10.1177/106480469500300403]
40. Christofidis MJ, Hill A, Horswill MS, Watson MO. Observation chart design features affect the detection of patient deterioration: a systematic experimental evaluation. J Adv Nurs 2016 Jan 09;72(1):158-172. [doi: 10.1111/jan.12824] [Medline: 26556775]

41. Preece MH, Hill A, Horswill MS, Karamatic R, Watson MO. Designing observation charts to optimize the detection of patient deterioriation: reliance on the subjective preferences of healthcare professionals is not enough. Aust Crit Care 2012 Nov;25(4):238-252. [doi: 10.1016/j.aucc.2012.01.003] [Medline: 22381946]

Abbreviations

HWD: head-worn display
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by L Buis; submitted 24.01.21; peer-reviewed by S Badawy, MD, MS, M Görges; comments to author 12.02.21; revised version received 26.02.21; accepted 30.03.21; published 11.05.21.

Please cite as:
Elrose F, Hill A, Liu D, Salisbury I, LeCong T, Loeb RG, Sanderson P
The Use of Head-Worn Displays for Vital Sign Monitoring in Critical and Acute Care: Systematic Review
JMIR Mhealth Uhealth 2021;9(5):e27165
URL: https://mhealth.jmir.org/2021/5/e27165
doi:10.2196/27165
PMID:33973863

Review

Use of Fitbit Devices in Physical Activity Intervention Studies Across the Life Course: Narrative Review

Ruth Gaelle St Fleur[1], BA; Sara Mijares St George[1], PhD; Rafael Leite[2], BA; Marissa Kobayashi[1], MHS; Yaray Agosto[1], MPH; Danielle E Jake-Schoffman[3], PhD

[1]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, United States

[2]Department of Psychology, University of Miami, Coral Gables, FL, United States

[3]Department of Health, Education, and Behavior, University of Florida, Gainesville, FL, United States

Corresponding Author:
Sara Mijares St George, PhD
Department of Public Health Sciences
University of Miami Miller School of Medicine
1120 NW 14th St
Miami, FL, 33136
United States
Phone: 1 305 2432000
Email: s.stgeorge@med.miami.edu

Abstract

Background: Commercial off-the-shelf activity trackers (eg, Fitbit) allow users to self-monitor their daily physical activity (PA), including the number of steps, type of PA, amount of sleep, and other features. Fitbits have been used as both measurement and intervention tools. However, it is not clear how they are being incorporated into PA intervention studies, and their use in specific age groups across the life course is not well understood.

Objective: This narrative review aims to characterize how PA intervention studies across the life course use Fitbit devices by synthesizing and summarizing information on device selection, intended use (intervention vs measurement tool), participant wear instructions, rates of adherence to device wear, strategies used to boost adherence, and the complementary use of other PA measures. This review provides intervention scientists with a synthesis of information that may inform future trials involving Fitbit devices.

Methods: We conducted a search of the Fitabase Fitbit Research Library, a database of studies published between 2012 and 2018. Of the 682 studies available on the Fitabase research library, 60 interventions met the eligibility criteria and were included in this review. A supplemental search in PubMed resulted in the inclusion of 15 additional articles published between 2019 and 2020. A total of 75 articles were reviewed, which represented interventions conducted in childhood; adolescence; and early, middle, and older adulthood.

Results: There was considerable heterogeneity in the use of Fitbit within and between developmental stages. Interventions for adults typically required longer wear periods, whereas studies on children and adolescents tended to have more limited device wear periods. Most studies used developmentally appropriate behavior change techniques and device wear instructions. Regardless of the developmental stage and intended Fitbit use (ie, measurement vs intervention tool), the most common strategies used to enhance wear time included sending participants reminders through texts or emails and asking participants to log their steps or synchronize their Fitbit data daily. The rates of adherence to the wear time criteria were reported using varying metrics. Most studies supplemented the use of Fitbit with additional objective or self-reported measures for PA.

Conclusions: Overall, the heterogeneity in Fitbit use across PA intervention studies reflects its relative novelty in the field of research. As the use of monitoring devices continues to expand in PA research, the lack of uniformity in study protocols and metrics of reported measures represents a major issue for comparability purposes. There is a need for increased transparency in the prospective registration of PA intervention studies. Researchers need to provide a clear rationale for the use of several PA measures and specify the source of their main PA outcome and how additional measures will be used in the context of Fitbit-based interventions.

(JMIR Mhealth Uhealth 2021;9(5):e23411) doi:10.2196/23411

KEYWORDS

physical activity; Fitbit; eHealth; life course; mobile phone

Introduction

Background

Insufficient physical activity (PA) in all stages of life, from early childhood to older adulthood, is a well-documented public health issue [1]. Between 2001 and 2016, although the levels of insufficient PA decreased marginally globally, high-income Western countries, such as the United States, reported a 5% increase in the prevalence of physical inactivity [2]. Insufficient PA is associated with increased risk for a variety of chronic diseases including cardiovascular disease, hypertension, and type 2 diabetes [3,4]. Although the current PA guidelines for Americans recommend at least 60 minutes per day of moderate- to vigorous-intensity PA for children and adolescents and 150 minutes per week of moderate-intensity PA for adults, more than 80% of youth and adults do not meet these guidelines [5].

Advances in 21st century technology have introduced the use of commercial off-the-shelf activity trackers (eg, Fitbit and Apple Watch) that allow users to self-monitor their daily PA. As one of the top 5 wearable companies based on shipment volume, Fitbit has produced some of the most popular fitness trackers that are currently available on the market [6]. These devices allow users to track their daily activities, including the number of steps, type of PA, and amount of sleep, among other features [7]. Fitbit released its first device in 2009 and its first wrist-worn tracker in 2012 [8]. The brand quickly gained popularity and saw a substantial increase in the use of activity trackers in a relatively short time. In 2014, Fitbit reported only 6.7 million active users compared with 29.6 million in 2019 [9]. In November 2019, Google announced its purchase of Fitbit for US $2.1 billion and publicly committed to accelerating innovation of these devices [7].

In the last decade, researchers have begun to take advantage of Fitbit's public appeal, prominence, and relatively low cost compared with that of other commercial off-the-shelf activity trackers such as the Apple Watch, by incorporating these devices into their studies. This has been facilitated by Fitbit's open application programming interface (API), which allows programmers to collect and store data across multiple devices [7]. Fitabase is an example of a company that capitalizes on Fitbit's open API and works with researchers to collect, manage, and analyze data from participants' Fitbit devices [10]. In addition to being a data management platform, Fitabase provides the general public with access to an extensive library containing hundreds of published studies, protocols, and methods papers that report their use of Fitbit devices [11]. As of January 7, 2021, 682 articles published between 2012 and 2018 were available on the Fitabase research library [11].

Objectives

Early studies involving Fitbit focused on establishing its accuracy as an objective PA measurement tool, especially in comparison with existing gold standard measurement devices [12,13]. The first study using a Fitbit device to assess PA was published in 2012 and assessed its validity in measuring steps taken during self-paced and prescribed PA [14]. Overall, there have been mixed findings about the accuracy of Fitbit measurements, with some studies indicating step count accuracy 50% of the time compared with research-grade accelerometers [15] and others reporting high validity in step count measurements [16,17]. In addition to their ability to serve as a PA measurement tool, Fitbit devices are increasingly being used to support self-monitoring and goal setting as a way of promoting PA in intervention studies across the life course [18-21]. However, it is not clear how these commercially available devices are being incorporated into PA intervention studies. This gap severely hinders the creation of standardized procedures that operationalize Fitbit use in PA intervention studies (eg, wear time protocols, strategies to boost wear time, and analysis implications) [22]. An overview of the ways in which Fitbit devices can be used to measure or help achieve the desired intervention effects can further contribute to the evidence base. Notably, Fitbit devices have been used in PA interventions targeting children through older adults. However, differences in use protocols across age groups (eg, models and strategies to boost wear time) are not known. In this context, this narrative review aims to characterize how PA intervention studies across the life course use Fitbit in terms of device selection, intended use (intervention vs measurement tool), wear instructions, rates of adherence to device wear, strategies used to boost adherence, and potential use of additional PA measures. This review provides intervention scientists with a synthesis of information that may inform future trials involving Fitbit devices.

Methods

Search Strategy and Eligibility Criteria

Given that it serves as a repository of Fitbit-related studies, we first conducted a search of the Fitabase Fitbit Research Library [11]. As of January 7, 2021, the Fitabase research library included studies published between 2012 and 2018 and retrieved from PubMed, Google Scholar, the Association for Computing Machinery, JMIR, Science Direct, and IEEE. Approximately twice a week during this period, the Fitabase team conducted searches of those sources using the keyword *Fitbit*. The studies identified in the search were then put through a screening process wherein they were deemed eligible for inclusion in the library only if a Fitbit device was used as a key element of the study (ie, for measurement or intervention purposes) [11]. In the Fitabase library, we applied preexisting filters to limit eligible studies to those that were (1) intervention studies, (2) focused on and reported PA as a main study outcome, and (3) conducted in one of five developmental stages of interest (ie, childhood [9-12 years old], adolescence [13-17 years old], early adulthood [18-40 years old], middle adulthood [41-64 years old], or older adulthood [≥65 years old]). We excluded nonintervention studies, those that did not report a specific target population, and those that did not have full-text articles available. We also excluded intervention studies that used Fitbit devices exclusively to monitor sleep. To capture studies

published between 2019 and 2020, we conducted a search of PubMed using the following string search: "(physical activity[Title/Abstract]) AND (Fitbit[Title/Abstract]) AND (intervention*[Title/Abstract])." In addition to applying the inclusion and exclusion criteria specified earlier, we excluded protocol and review papers and qualitative studies.

Data Collection

The first 2 authors created a standardized form for data extraction by using Microsoft Excel. The items on this form, which were all open-ended, captured (1) general study characteristics (ie, sample size, study design, and intervention description) and (2) Fitbit use (ie, model, wear time and adherence, strategies to boost wear time, and other measures of PA). After finalizing the form, the first author read all the eligible studies and extracted the relevant data. To enhance the reliability of the extracted information, 3 additional coders (RL, MK, and YA) subsequently read the articles and reviewed the extracted data. As part of our protocol, disagreements between authors were resolved through discussion, with the final decision being made by the senior author.

Results

Overview

Of the 682 studies available on the Fitabase Fitbit Research Library, 60 interventions met the eligibility criteria for this review. An additional 15 eligible studies resulting from the PubMed search were included. A total of 75 studies were reviewed (n=6 in childhood, n=11 in adolescence, n=20 in early adulthood, n=28 in middle adulthood, and n=10 in older adulthood). Figure 1 shows the flow diagram of the study. Tables 1 and 2 show the study characteristics and Fitbit use by developmental stage for included studies, organized by intended Fitbit use (ie, intervention vs measurement).

Figure 1. Study selection flow diagram.

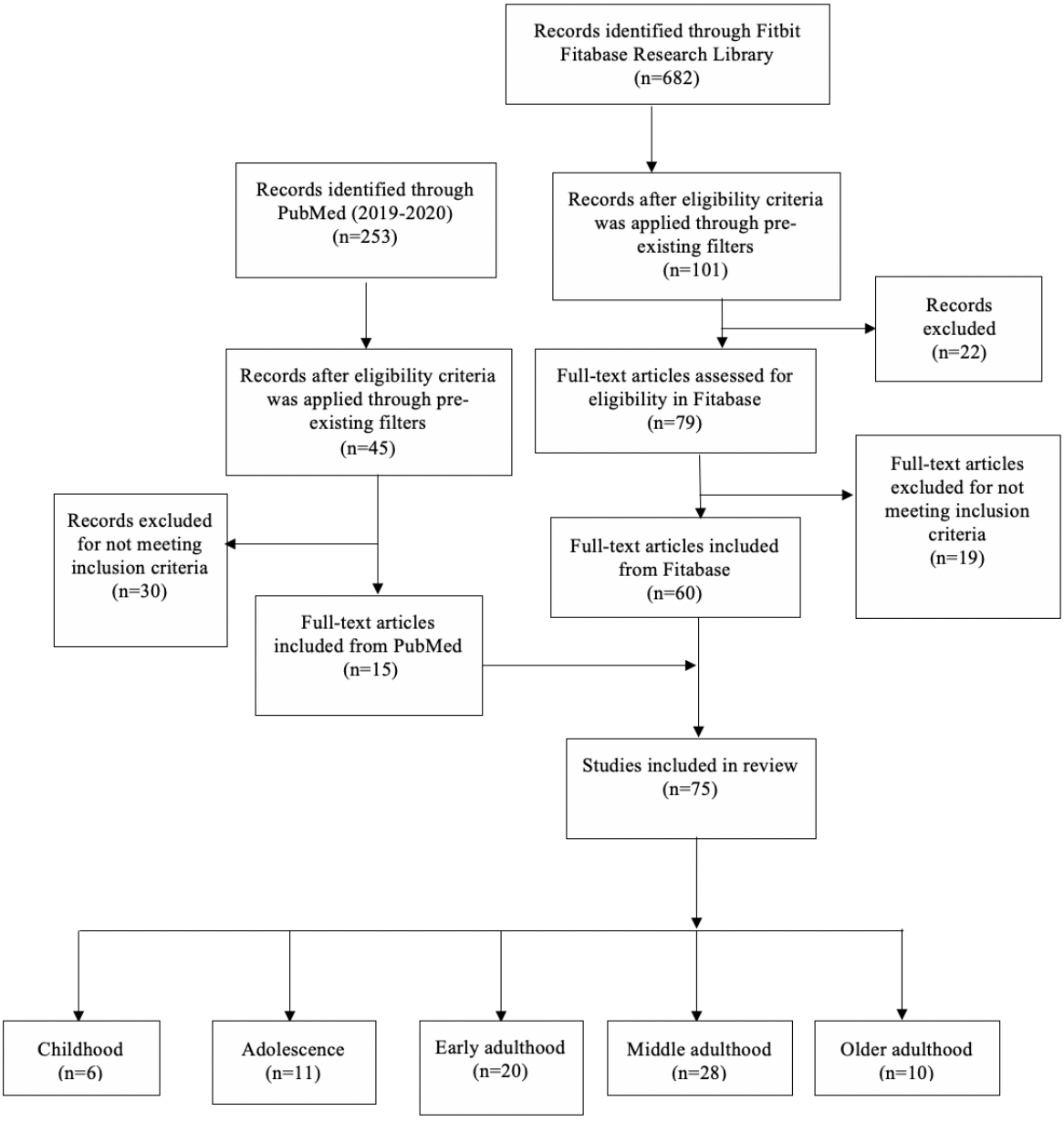


Table 1. General study characteristics.

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Childhood						
Intervention only						
Evans et al, 2017 [23]	• Quasi-experimental design with 3 conditions: (1) Fitbit+intervention, (2) Fitbit only, and (3) control • 6-week classroom-based intervention • One session per week lasting 40 min and led by teachers and study staff • Individual and group-level achievements • BCTs[a]: goal setting, self-monitoring, and rewards	42	12.3 (0.3)	47[b]	NR[c]	42% overweight or obese
Mackintosh et al, 2016 [24]	• Single-group pre-post design • 4-week intervention with teams designing and completing week-long missions • Teachers equipped with a guide and DVD outlining various missions • BCTs: goal setting, self-monitoring, and rewards	30	10.1 (0.3)	40	NR	BMI: mean 19.9 (SD 4) kg/m^2
Measurement only						
Walther et al, 2018 [25]	• Single-group pre-post design • 12-week afterschool program with two 60-min sessions per week (24 total) • 12 sessions focused on nutrition and increasing PA[d] and 12 sessions taught safe food preparation while preparing simple, healthful recipes • BCTs: shaping knowledge and self-monitoring	24	9.58 (NR)	83	30% White; 29% Black; 25% Hispanic; 16% Native American	NR
Intervention and measurement						
Buchele Harris and Chen, 2018 [18]	• Quasi-experimental design with 2 conditions: (1) PA engaging the brain+Fitbit challenge (PAEB-C) or (2) Fitbit only • 4-week school-based intervention • Participants in PAEB-C condition followed a 6-min video once a day • BCTs: behavioral rehearsal and self-monitoring	116	10-11	49	60% reported race other than White, with 30% Black[b]	NR
Harris et al, 2018[b] [26]	• Quasi-experimental design with 2 conditions: (1) coordinated-bilateral PA intervention or (2) Fitbit only • 4-week school-based intervention • Repetitive coordinated-bilateral motor movements performed while following a 6-min video instruction once a day • BCTs: behavioral rehearsal and self-monitoring	116	NR	50	60% reported race other than White, with 30% Black[b]	NR
Hayes and Van Camp, 2015 [27]	• Single-group pre-post design • 22 sessions of 20 min, 1 to 4 days per week on an elementary school playground during regularly scheduled, unstructured recess • BCTs: self-monitoring	6	NR	100	NR	66% normal weight
Adolescence						
Intervention only						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Chen et al, 2017 [28]	• RCT[e] with 2 conditions • Phone-based 3-month intervention for adolescents who are overweight and obese • 8 modules focused on lifestyle modification, weight management, nutrition, and stress • BCTs: shaping knowledge and self-monitoring	40	14.9 (1.7)	42	90% Chinese American	BMI: mean 28.3 (SD 4.7) kg/m^2
Gandrud et al, 2018 [29]	• Parallel-group RCT with 2 conditions • 6-month intervention using intensive remote therapy for pediatric patients with type 1 diabetes • Content focused on recommendations for diabetes management, glucose control, and PA • BCT: shaping knowledge and self-monitoring	117	12.7 (2.5)	54	NR	BMI z-score: mean 0.5 (SD 0.9)
Mendoza et al, 2017 [30]	• Pilot RCT with 2 conditions • 10-week intervention for adolescent and young adult survivors of cancer using a wearable device, mobile health app, and Facebook support group for reaching PA goals • BCTs: shaping knowledge, self-monitoring, and social support	60	16.6 (1.5)	59	66% non-Hispanic White; 14% Hispanic; 7% non-Hispanic Black; 14% Other	NR
Measurement only						
Haegele and Porretta, 2016 [31]	• Single-group pre-post design • Social cognitive theory–based PA intervention for adolescents with visual impairments • 9 lessons delivered during PA classes that included curricular concepts, in-class activities, and homework • BCTs: shaping knowledge, behavioral rehearsal, and self-monitoring	6	NR	NR	NR	NR
Meng et al, 2018 [32]	• Quasi-experimental design • 2-year intervention for soccer players delivered by coaches • Content focused on addressing exercise, body image, and nutrition • BCTs: shaping knowledge and self-monitoring	388	15.3 (1.1)	58	62% non-Latino; 38% Latino	BMI %: mean 62.8 (SD 25.0)
Walther et al, 2018 [25]	• Pre-post study design • 12-week intervention with fourth and fifth graders that focused on proper nutrition and safe food preparation techniques and promoted PA via interactive games • BCTs: self-monitoring, shaping knowledge, and social support	30	9.58 (NR)	83	30% White; 29% Black or African American; 25% Hispanic; 16% Native American	NR
Intervention and measurement						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Gaudet et al, 2017 [19]	• Quasi-experimental crossover design • 7-week classroom-based intervention to increase students' PA • BCTs: self-monitoring, self-regulation, and goal setting	46	13.0 (0.3)	52%	NR	NR
Pope et al, 2018 [33]	• Multiphase mixed methods consisting of an RCT • 12-week intervention for high school students where participants assigned to the game group were rewarded based on the number of daily steps taken • BCTs: goal setting, self-monitoring, and rewards	105	17.0 (NR)	71	67% White; 16% Black; 12% Hispanic or Latino; 12% Asian; 5% Other	NR
Remmert et al, 2019 [34]	• Quasi-experimental pilot study • 12-week school-based ABT[f] intervention to increase PA in adolescents with low activity • Weekly sessions conducted by project coordinator consisted of acceptance-based behavioral counseling combined with preferred-intensity exercise for 30 min • BCTs: behavioral counseling, behavioral practice, and self-monitoring	20	12.0 (0.0)	60	55% Latino; 25% non-Latino White; 20% Other	BMI: mean 21.7 (SD 3.6) kg/m^2
Short et al, 2018 [35]	• RCT with 2 conditions • 48-week exercise intervention subdivided into 3 consecutive 16-week phases • Tested how different incentive schemes influence exercise frequency and duration among youth • Self-monitoring and rewards	77	14.0 (2.2)	NR	100% American Indian	BMI%: mean 98 (SD 3)
Van Woudenberg et al, 2018 [36]	• RCT with 2 conditions • 7-day classroom-based intervention that used a social network model to select and train influential adolescents (using smartphones) • BCTs: social facilitation, behavior modeling, impression management, and self-persuasion	190	12.2 (0.5)	54	NR	NR
Early adulthood (18-40 years)						
Intervention only						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Bang et al, 2017 [37]	• Quasi-experimental design • 6-week campus-based program with one session per week during lunch • Participants walked together through the campus forest for approximately 40 min and received one lecture on stress management • Encouraged to walk at least once per week at their leisure • BCTs: self-monitoring, behavioral practice, and social support	99	24.8 (4.7)[b]	49[b]	NR	BMI: mean 21.9 (SD 2.9) kg/m^{2}[b]
Baruth et al, 2019 [38]	• Quasi-experimental pilot study with 2 conditions: (1) intervention and (2) control • Weekly PA intervention for pregnant women until 35-week gestation • BCTs: goal setting, behavior counseling, self-monitoring, and social support	45	28.4 (4.5)[b]	100	81.8% White[b]	BMI: mean 26.9 (SD 7.2) kg/m^{2}[b]
Losina et al, 2017 [39]	• Single condition feasibility study • 6-month workplace program to increase PA among sedentary hospital employees through individual and team-based financial incentives • BCTs: self-monitoring, goal setting, and rewards	292	38.0 (11.0)	83	62% White; 14% Black; 10% Asian; 7% Hispanic; 7% Other	32% normal weight; 30% overweight; 38% obese
Mahar et al, 2015 [40]	• RCT with 2 conditions: (1) Fitbit and (2) no Fitbit • 10-week intervention examined effects of movement technology on college students' PA • BCTs: self-monitoring	75	19.4 (1.2)	NR	NR	NR
Measurement only						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Chen and Pu, 2014 [41]	• RCT with 3 conditions: (1) competition, (2) cooperation or (3) hybrid • One-week mobile app intervention to help promote exercise in pairs and earn badges based on performance • BCTs: self-monitoring, social support, goal setting, and rewards	36	20-30	58	NR	2.8% underweight, 94% normal weight, 2.8% obese
Pagkalos et al, 2017 [42]	• RCT with 2 conditions: (1) intervention and (2) control • 5-week pilot study to monitor young adults' exercise via a custom-built Facebook app for activity self-reporting • BCTs: self-monitoring and social support	49	24.0 (7.0)	NR	NR	BMI: mean 22.5 (SD 3.0) kg/m^2
Ptomey et al, 2018 [43]	• RCT with 2 conditions: (1) exercise once a week and (2) exercise twice a week • 12-week at-home intervention to increase MVPA[g] using videoconferencing for groups of adults with Down syndrome • BCTs: self-monitoring, behavioral practice, and social support	27	27.9 (7.1)	41	10% ethnic minorities	Group 1 BMI: mean 35.4 (SD 9.7) kg/m^2; Group 2 BMI: mean 31.4 (SD 6.8) kg/m^2
Walsh and Golbeck, 2014 [44]	• Within-subject crossover study with 3 conditions: (1) social game using Fitbit steps as currency, (2) social interaction experience, and (3) control • 30-day web-based intervention • Participants in the social interaction could interact or communicate and share their PA levels with friends • BCTs: self-monitoring, social support, and social comparison	74	37.7 (10.2)	59	NR	NR
Yoon et al, 2018 [45]	• RCT with 2 conditions: (1) intervention and (2) control • Observational PA data collected from participants over first 6 months • Participants were sent a personalized email message about their activity to inform them of current PA levels and encourage increase in the last 6 months • BCTs: self-monitoring and feedback on behavior	79	31.9 (9.6)	59	29.2% Hispanic	NR
Intervention and measurement						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Choi, 2016 [46]	• RCT with 2 conditions: (1) intervention mobile app+Fitbit and (2) Fitbit • 12-week intervention with pregnant women between 10 and 20 weeks of gestation • After an initial 30-min in-person intervention session, participants received daily message or video, encouragement, and activity diary through the app • BCTs: self-monitoring, shaping knowledge, and written persuasion to boost self-efficacy	30	33.7 (2.6)	100	43% White; 40% Asian; 10% Hispanic; 7% Black	BMI (prepregnancy): mean 27.7 (SD 3.7) kg/m^2
Chung et al, 2017 [47]	• Single-group pre-post design stratified into 2 groups: (1) overweight or obese group and (2) healthy weight group • 2-month intervention where participants received Twitter messages to encourage PA and healthy eating, photo-based messages, infographics, and website links related to healthy lifestyle behaviors • BCTs: self-monitoring, shaping knowledge, and written persuasion to boost self-efficacy	12	19-20	67	50% White; 33% Black; 8% Asian; 8% American Indian	Group 1 BMI range: 25-35 kg/m^2; Group 2 BMI range: 22-24.9 kg/m^2
Gilmore et al, 2017 [48]	• RCT for postpartum women with 2 conditions: (1) WIC[h] standard care (WIC Moms) and (2) WIC standard care and personalized weight management via a smartphone (E-Moms) • E-Moms group was given access to the SmartLoss SmartPhone app that included near real-time weight and activity monitoring, scheduled delivery of health information, and interventionist feedback • BCTs: self-monitoring, feedback on behavior	35	26.0 (5.4)	100	74% African American	BMI: mean 32 (SD 3) kg/m^2 (range 25.6-37.0 kg/m^2)
Halliday et al, 2017 [49]	• Pre-post study design • A goal-focused exercise program that included weekly phone or face-to-face coaching to reinforce walking goals, as well as an optional 1-h supervised group walk on 2 occasions per week • BCTs: self-monitoring, social support, behavioral practice, behavior counseling, goal setting	15	38.3 (6.4)	60	80% Caucasian	BMI: mean 30.4 (SD 6.4) kg/m^2
Florence et al, 2016 [50]	• RCT with 3 conditions: (1) group 1 (Fitbit+modules), (2) group 2 (Fitbit+modules+a social media-based game), (3) control group with just educational modules • 14-week intervention for first-year medical students where daily steps and sleep hours were monitored in groups 1 and 2 during weeks 1-8 • From week 9, all 3 groups had access to Fitbit Flex and the game platform, and students' daily steps and sleep time were monitored until week 14 by Fitbit Flex • BCTs: self-monitoring and social support	300	18-19	58	NR	NR
Miragall et al, 2017 [51]	• RCT with 3 conditions: (1) IMI[i]+PED condition (access to IMI and use of a pedometer), (2) IMI condition (access to IMI and use of a blinded pedometer), and (3) control condition (use of a blinded pedometer) • 3-week IMI conducted with sedentary or low-active students to increase motivation and set individualized PA goals • BCTs: self-monitoring, goal setting, and verbal persuasion about self-efficacy	76	22.2 (3.7)	86	NR	BMI: mean 21.7 (SD 3.2) kg/m^2

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Schrager et al, 2017 [52]	• Pre-post cohort study • 1-month intervention where emergency medicine residents were asked to wear a Fitbit to assess its effects on their PA levels • BCTs: self-monitoring	30	Median age: 28	47	NR	NR
Thorndike et al, 2014 [53]	• 2-phase intervention: phase 1 was a 6-week RCT and phase 2 was a 6-week nonrandomized team steps competition • 12-week intervention that provided medical residents with free access to a fitness center, weekly one-hour personal training sessions, and up to 2 individual appointments with a Be Fit staff nutritionist • BCTs: self-monitoring and shaping knowledge	108	29 (23-37)	54	66% White	BMI: mean 24.1 (range 17.8-35.6) kg/m^2
Washington et al, 2014 [54]	• Pre-post study design • 3-week intervention in which participants won prizes for wearing their Fitbit and meeting experimenter-determined step criteria • BCTs: self-monitoring, goal setting, and rewards	13	18-26	67	NR	NR
West et al, 2016 [55]	• Quasi-experimental study design • 9-week intervention where undergraduate students were assigned to either (1) a behavioral weight gain prevention intervention (healthy weight) or (2) an HPV[j] awareness intervention • 8 lessons on behavioral strategies to maintain weight and avoid obesity were delivered via electronic newsletters and Facebook postings • BCTs: self-monitoring and shaping knowledge	58	21.6 (2.2)	81	90% White	BMI: mean 24.0 (SD 5.1) kg/m^2
Zhang and Jemmott, 2019 [56]	• Pilot RCT with 2 conditions: (1) intervention and (2) control • 3-month intervention in small groups with mobile app to track group's PA data and engage with others • BCTs: self-monitoring, social support, and social comparison	91	26.8 (5.1)	100	100% African American	BMI: mean 31.6 (SD 8.2) kg/m^2
Middle adulthood (41-64 years)						
Intervention only						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Amorim et al, 2019 [57]	• Pilot RCT with 2 conditions: (1) intervention and (2) control • 6-month intervention with PA booklet, health coaching sessions, app, and Fitbit • BCTs: self-monitoring, behavioral counseling, and shaping knowledge	68	58.4 (13.4)	50	NR	BMI: mean 28 (SD 5.5) kg/m^2
Butryn et al, 2014 [58]	• Single-group pre-post design • 6 months group-based intervention with a web platform component to facilitate social connectivity • BCTs: self-monitoring and social support	36	54 (7.18)	100	62% Caucasian	BMI: mean 32.7 (SD 7.32) kg/m^2
Cadmus-Bertram al et, 2015 [59]	• RCT with 2 conditions: (1) intervention (2) comparison (standard pedometer only) • 16-week web-based self-monitoring intervention for inactive, postmenopausal women • Content combined self-monitoring with self-regulatory skills, such as goal setting and frequent feedback • BCTs: self-monitoring, knowledge shaping, self-regulation, goal setting, and feedback	51	60.0 (7.1)	100	92% non-Hispanic White[b]	BMI: mean 29.2 (SD 3.5) kg/m^2
Cadmus-Bertram et al, 2019 [60]	• Pilot RCT with 2 conditions: (1) intervention and (2) comparison • 12-week multi-component intervention for cancer survivors and support partners with Fitbit linked to electronic health records • BCTs: self-monitoring and social support	50	54.4 (11.2)	96	94% non-Hispanic White; 2% Hispanic; 2% Black; 2% Multiracial	BMI: mean 32.2 (SD 7.4) kg/m^2
Dean et al, 2018 [20]	• Quasi-experimental pilot study • 8 weekly small group sessions • Each 90-min session had a group discussion and an exercise component • BCTs: self-monitoring, knowledge shaping, and social support	40	46.9 (9.8)	0	100% African American	67% obese
Duncan et al, 2020 [61]	• RCT with 3 conditions: (1) enhanced, (2) traditional, and (3) control • 6-month intervention for adults with overweight or obesity delivered via the app with educational content, dietary consultation, Fitbit, and scales • Enhanced group received additional sleep intervention content via the app • BCTs: self-monitoring, knowledge shaping, goal setting, and behavioral counseling	116	44.5 (10.5)	70.7	NR	BMI: mean 31.7 (SD 3.9) kg/m^2
Ellingson et al, 2019 [62]	• Randomized feasibility trial with 2 conditions: (1) intervention with Fitbit and (2) Fitbit only • 12-week intervention with motivational interviewing, habit education, and Fitbit • BCTs: self-monitoring and verbal persuasion to boost self-efficacy	91	41.7 (9.3)	53	79% White	BMI: mean 29.6 (SD 6.3) kg/m^2
Kandula et al, 2017 [63]	• 16-week community-based, pre-post intervention • Twice weekly group exercise classes, Fitbit Zip and web-based platform, goal setting, and classes on healthy eating • BCTs: self-monitoring, social support, goal setting, and knowledge shaping	30	40 (5)	100	100% South Asian	BMI: mean 30 (SD 3) kg/m^2
		80	51.1 (11.7)	86		

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Ross and Wing, 2016 [64]	• Randomized pilot trial with 3 conditions: (1) tech, (2) tech+phone, and (3) self-monitoring • 6-month intervention with one group receiving self-monitoring tools (eg, booklets or scale) • Tech group received Fitbit and tracked caloric intake through Fitbit app • Tech+phone group received same materials along with 14 calls regarding behavioral weight loss techniques • BCTs: self-monitoring, behavioral counseling, and knowledge shaping				84% Non-Hispanic White	BMI: mean 33 (SD 3.4) kg/m^2
Singh et al, 2020 [65]	• RCT with 2 conditions: (1) PA counseling, (2) PA counseling and Fitbit • 12-week intervention for women with breast cancer that included a PA counseling session with exercise physiologist and educational booklet • BCTs: self-monitoring, behavioral counseling, and knowledge shaping	52	Group 1: 52.8 (9.5); Group 2: 49.5 (8.6)	100	NR	Group 1: BMI: mean 28.5 (SD 5.2) kg/m^2; Group 2: BMI: mean 28.7 (SD 6) kg/m^2
Van Blarigan et al, 2019 [66]	• Pilot RCT with 2 conditions: (1) intervention and (2) control • 12-week intervention for cancer survivors with daily text messaging • BCTs: self-monitoring and cues	42	54 (11)	59	73% White, 12% Asian, 12% Native American or other, 2% Black	BMI: mean 28.4 (SD 5.9) kg/m^2
Measurement only						
Patel et al, 2017 [67]	• 12-week family-based RCT intervention • On the basis of behavioral economics and gamification principles, the intervention used points and levels (bronze, silver, gold, and platinum) to encourage families to change their behavior and increase their PA levels • BCTs: self-monitoring, rewards, and social support	200	55.4 (NR)	56	100% Caucasian	BMI: mean 27.2 (SD 5.1) kg/m^{2b}
Robinson et al, 2019 [68]	• Pilot RCT with 2 conditions: (1) intervention and (2) control • 5-week study using implementation intentions to establish PA habits using personalized materials • BCTs: self-monitoring and knowledge shaping	63	49.4 (8.3)	72.6	NR	NR
Schumacher et al, 2017 [69]	• Single-group pre-post trial study • Partner-based PA program for women examining PA lapses, cognitive-affective responses to lapses, and the role of social support in PA • BCTs: self-monitoring and social support	20	50 (7.2)	100	95% Caucasian	BMI: mean 30.9 (SD 8.9) kg/m^2
Intervention and measurement						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Adams et al, 2017 [70]	• 2×2 factorial, 4-month RCT with goal setting (adaptive vs static goals) and rewards (immediate vs delayed) • WalkIT trial delivered intervention components by SMS text messages on a daily basis with prompt-to-action messages (eg, tips, questions, or motivational or inspirational messages) • BCTs: self-monitoring, goal setting, shaping knowledge, persuasion to boost self-efficacy, and cues	96	41 (9.5)	77	81.3% Caucasian	BMI: mean 34.1 (SD 6.18) kg/m^2
Arigo, 2015 [71]	• Single-group pre-post design • 4-week web-based intervention in pairs • Participants have access to web-based modules and worksheets guiding them through seeking support and setting weekly PA goals • BCTs: self-monitoring, social support, and goal setting	12	46 (13.1)	100	75% Caucasian	BMI: mean 32.6 (SD 5.7) kg/m^2
Arigo et al, 2015b [72]	• Single-group pre-post design • 6-week program predominantly web-based with a single face-to-face session introducing PA promotion skills • Participants were encouraged to communicate with their PA dyad partner and other participants • BCTs: self-monitoring, goal setting, and social support	20	50 (7.2)	100	90% Caucasian	BMI: mean 30.9 (SD 8.9) kg/m^2
Finkelstein et al, 2015 [73]	• Randomized crossover design with 2 conditions: (1) message-on and (2) message-off • 4-week web-based intervention targeted inactivity level with tailored text messages about sedentary time • BCTs: self-monitoring and cues	27	52 (12.0)	100	47% White; 47% African American	BMI: mean 37.0 (SD 6.0) kg/m^2
Fukuoka et al, 2018 [74]	• Single-group pre-post trial, uncontrolled pilot study • 8-week weight loss program for Latino adults at risk for type 2 diabetes • Participants were provided with 2 in-person counseling sessions, Fitbit, use of the Fitbit app, and a Facebook group and were asked to track diet daily and weight twice per week • BCTs: self-monitoring, behavioral practice, and social support	54	45.3 (10.8)	68.5	100% Latino	BMI: mean 31.4 (SD 4.1) kg/m^2
Gell et al, 2020 [75]	• Pilot RCT with 2 conditions: (1) intervention and (2) control with Fitbit • 8-week intervention for cancer survivors with health coaching, text messaging, and Fitbit • BCTs: self-monitoring, behavioral counseling, and cues	59	61.4 (9)	81	98.5% non-Hispanic White, 1.2% Black or Hispanic	BMI: mean 30.4 (SD 7) kg/m^2
Gremaud et al, 2018 [76]	• 10-week RCT intervention comparing 2 arms: (1) Fitbit only and (2) Fitbit+MapTrek • MapTrek, mobile phone–based walking game leverages Fitbit to track users' PA and motivate users to engage in virtual walking races in numerous places around the globe • BCTs: self-monitoring and feedback	146	40.6 (11.7)[b]	79.2[b]	91.7% Caucasian[b]	BMI: mean 29.9 (SD 6.6) kg/m^{2b}
		11	59.53 (11.7)	100	NR	

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Grossman et al, 2017 [77]	• 16-week behavioral pre-post pilot program for postmenopausal women • The program consisted of face-to-face group meetings every month, weekly weigh-ins, electronic check-ins, calorie-restricted diet, and high-intensity interval training • BCTs: self-monitoring, social support, and behavioral practice					BMI: mean 32 (SD 2.53) kg/m^2
Linke et al, 2019 [78]	• One-arm pilot study • 12-week intervention for veterans recovering from substance use disorder that included psychoeducation classes, gym membership, and Fitbit • BCTs: self-monitoring, social support, and knowledge shaping	15	45 (9.7)	13	60% non-Hispanic White, 27% Black, 13% Hispanic	NR
Meints et al, 2019 [79]	• Prospective cohort study • 26-week intervention for hospital employees to increase PA with financial incentives • Groups of 3 were formed and financial incentives were given if team members met goals • BCTs: self-monitoring, social support, rewards, and goal setting	225	Black participants: 43 (10); White participants: 39 (12)	84	81% White; 19% Black	Black participants: 84% had overweight or obesity; White participants: 68% had overweight or obesity
Painter et al, 2017 [80]	• Retrospective analyses of 6 weight loss programs • Participants were taught self-management strategies and were given a Fitbit, Wi-Fi-enabled scale, digital food and exercise log, and access to expert coach via electronic messages • BCTs: self-monitoring and behavioral counseling	2113	44.54 (10.72)	59	NR	BMI: mean 33.8 (SD 6.8) kg/m^2
Reed et al, 2019 [81]	• Randomized repeated-measures study with 2 conditions: (1) intervention and (2) control • 12-week intervention with self-regulatory PA strategies, weekly text messaging, and Fitbit • BCTs: self-monitoring, self-regulation, and cues	59	48 (NR)	79.3[b]	93.2% White[b]	Weight: mean 92.47 (SD 22.8) kg[b]
Wang et al, 2015 [82]	• RCT with 2 conditions: (1) text messaging+Fitbit and (2) Fitbit only • 6-week intervention for adults with overweight and obesity receiving Fitbit and 3 daily SMS text messages prompting PA • BCTs: self-monitoring and cues	67	48.2 (11.7)	91	67% White; 16% Hispanic; 4% African American; 3% Asian; 3% Other	BMI: mean 31 (SD 3.7) kg/m^2
Willis et al, 2017 [83]	• Randomized feasibility study with 2 conditions: (1) web-based social network delivery and (2) conference call delivery • 6-month weight loss intervention • Web-based social network condition had 24 weekly web-based modules led by health educators • Conference call condition consisted of 24 weekly 60-min phone conferences • BCTs: self-monitoring, social support, and knowledge shaping	70	47 (12.4)	84	24.3% minorities	BMI: mean 36.2 (SD 4) kg/m^2
Older adulthood (≥65 years)						
Intervention only						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Ashe et al, 2015 [84]	• Randomized pilot trial with 2 conditions: (1) intervention and (2) comparison (educational sessions) • 6-month intervention to increase PA through social support, group-based education, and individualized PA prescription • BCTs: self-monitoring, knowledge shaping, and social support	25	64.1 (4.6)	100	NR	BMI: mean 26.9 (SD 6.8) kg/m^{2b}
Christiansen et al, 2020 [85]	• RCT with 2 conditions: (1) intervention and (2) control • 6-month intervention for total knee replacement patients that included physical therapy, Fitbit, step goals, and monthly call with physical therapist • BCTs: self-monitoring, goal setting, and behavioral counseling	43	67 (7)	53.4	91% White	BMI: mean 31.5 (SD 5.9) kg/m^2
Kenfield et al, 2019 [86]	• Pilot RCT with 2 conditions: (1) intervention and (2) control • 12-week intervention for men with prostate cancer that included personalized health recommendations, Fitbit, study website, and text messages • BCTs: self-monitoring, knowledge shaping and cues	76	65 (NR)	0	84% White	41% overweight, 35% with obesity
Thompson et al, 2014 [21]	• Randomized controlled crossover trial with 2 conditions: (1) immediate intervention and (2) delayed intervention • 48-week total: 24-week intervention that combined accelerometers with exercise counseling and 24 weeks without intervention • Content included materials on exercise, goal setting, and tracking PA • BCTs: self-monitoring, goal setting, behavioral counseling, and knowledge shaping	48	79.5 (7.0)	81	NR	Weight: mean 75.7 (SD 13.4) kg^b
Measurement only						
Rossi et al, 2018 [87]	• Single-group study (survey and qualitative interviews) • Participants wore Fitbit for 30 days to evaluate acceptability and validity of the device in diverse cancer survivors • BCTs: self-monitoring	25	62 (9)	100	36% non-Hispanic White; 36% Hispanic; 16% non-Hispanic Black; 12% Asian	BMI: mean 32 (SD 9) kg/m^2
Schmidt et al, 2018 [88]	• Single-group study • Participants wore Fitbit for 14 consecutive days and social cognitive factors, health issues, and views on aging were assessed • BCTs: self-monitoring	40	66.3 (3.19)	62.5	NR	BMI: mean 25.19 (SD 3.52) kg/m^2
Streber et al, 2017 [89]	• RCT with 2 conditions: (1) intervention and (2) control with weekly gymnastics or cognitive training • 12-week intervention with 90-min weekly sessions including PA program with social and cognitive activities and PA coaching program • BCTs: self-monitoring, social support, knowledge shaping, and behavioral counseling	87	76 (9.2)	78	NR	NR
Intervention and measurement						

Developmental stage	Study design and intervention description	Participant characteristics at baseline				
		Value, N	Age (years), mean (SD) or range	Female, %	Race or ethnicity	Weight status (eg, BMI, weight)
Harkins et al, 2017 [90]	• RCT with 4 conditions: (1) financial incentive, (2) social goals, (3) combined, and (4) control • 16-week intervention to test use of financial incentives and donations on PA increase with 4-week follow-up that included pedometer, goal setting, and weekly feedback on goal attainment • BCTs: self-monitoring, rewards, goal setting, and feedback	94	80.3	74	98% Caucasian	NR
McMahon et al, 2017 [91]	• 2×2 randomized factorial experiment with 4 conditions receiving PA protocol and Fitbit: (1) interpersonal BCS[k], (2) intrapersonal BCS, (3) interpersonal and intrapersonal BCS, and (4) control based on receipt of interpersonal and intrapersonal behavior change strategies • 8-week intervention with weekly 90-min meetings with all conditions receiving PA protocol, Fitbit, and workbook • BCTs: self-monitoring, knowledge shaping, and social support	102	79 (NR)	75	75% White; 25% Black	NR
Vidoni et al, 2016 [92]	• Randomized crossover trial with 2 conditions: (1) immediate intervention and (2) delayed intervention • 16-week trial divided into 8-week intervention and 8-week baseline or maintenance phase data collection • Intervention included the use of a Fitbit device and PA prescription • BCTs: self-monitoring and goal setting	30	With cognitive impairment: 72.3 (5.2); without cognitive impairment: 69.6 (5.8)	With cognitive impairment: 43; without cognitive impairment: 89	With cognitive impairment: 90% White; 10% African- American; without cognitive impairment: 100% White	BMI (with cognitive impairment): mean 29.4 (SD 3.8) kg/m^2; BMI (without cognitive impairment): mean 27.8 (SD 4.3) kg/m^2

[a]BCT: behavior change technique.

[b]Only intervention condition data reported.

[c]NR: not reported.

[d]PA: physical activity.

[e]RCT: randomized controlled trial.

[f]ABT: acceptance-based therapy.

[g]MVPA: moderate-to-vigorous physical activity.

[h]WIC: women, infants, and children.

[i]IMI: internet-based motivational intervention.

[j]HPV: human papillomavirus.

[k]BCS: behavior change strategy.

Table 2. Description of Fitbit use.

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Childhood							
Intervention only							
Evans et al, 2017 [23]	Zip (phase 1) and charge (phase 2)	Phase 1: all waking hours 7 days/week; phase 2: 24 h, 7 days/week	Minimum of 8 h/day	Days participants were adherent in phase 1: 64.8%; days participants were adherent in phase 2: 73.4%[b]	After-session meetings with study staff to sync their Fitbit data	Yes; same for Fitbit-only comparison condition; no device for control group	Sensewear, Armband Mini, and Jawbone
Mackintosh et al, 2016 [24]	Zip	Duration of intervention	Entire duration of session	100% adherence (with staff monitoring)	NR[c]	N/A[d]	Accelerometry
Measurement only							
Walther et al, 2018 [25]	Charge HR	24 h for 7 days, including one weekend	NR	NR	NR	N/A	Self-reporting
Intervention and measurement							
Buchele Harris and Chen, 2018 [18]	Charge HR	Daily; 5 school days/week for 4 weeks	Minimum of 14 h/day	Average loss of 1-day data per person per week	Log sheets record PA	No	NR
Harris et al, 2018[b] [26]	Charge HR	Daily; 5 school days/week for 4 weeks	NR	NR	Devices were charged at the end of the week	Yes; same use	NR
Hayes and Van Camp, 2015 [27]	Classic	Duration of intervention recess session	Entire duration of 20-min recess session	100% adherence (with staff monitoring)	NR	N/A	Second Fitbit
Adolescence							
Intervention only							
Chen et al, 2017 [28]	Flex	Daily for 3 months	NR	NR	Weekly text reminders and phone calls	No	Self-reporting of PA using the California Health Interview Survey
Gandrud et al, 2018 [29]	NR	NR	NR	NR	Weekly reminders sent to upload data	Yes	NR
Mendoza et al, 2017 [30]	Flex	Daily for 10 weeks	Minimum of 500 steps/day	Days participants were adherent: 72%	Text reminders sent every other day to encourage PA goals	No	Accelerometry
Measurement only							
Haegele and Porretta, 2016 [31]	Zip	NR	NR	NR	NR	N/A	NR
Meng et al, 2018 [32]	Zip	7 days/week at baseline and post measures	Minimum of 8 h/day	NR	Daily texts or email reminders	Yes; device masked with duct-tape	NR

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Walther et al, 2018 [25]	Charge	Wear on the 2nd and 10th week of the intervention for 7 days, including 1 weekend	24 h	NR	NR	N/A	Self-reported days of 60-min PA
Intervention and measurement							
Gaudet et al, 2017 [19]	Charge HR	Daily for 7 weeks	Minimum of 10 h/day	Median participant adherent 67% of intervention days	NR	Yes	Accelerometry and self-reporting
Pope et al, 2018 [33]	Flex	Daily for 12 weeks	NR	15% of students wore their Fitbit for <10 days; 36% never wore their Fitbit	Weekly lottery to win US $10 Amazon gift cards, weekly email reminders, and in-person troubleshooting at school once a week	Yes	NR
Remmert et al, 2019 [34]	Flex 2	Daily for 12 weeks	NR	Average number of days of valid Fitbit wear: 78 (out of 84 days)[b]	NR	Yes	Accelerometry
Short et al, 2018 [35]	Zip	Daily for 7 days	NR	NR	NR	Yes	NR
Van Woudenberg et al, 2018 [36]	Flex	Daily for 7 days	Minimum of 1000 steps/day	Days participants were adherent: 73.4%	NR	Yes	NR
Early adulthood (18-40 years)							
Intervention only							
Bang et al, 2017 [37]	Zip	NR	NR	NR	NR	No	IPAQ[e]
Baruth et al, 2019 [38]	Charge	Daily for duration of intervention	Minimum one day per week	Fitbit worn on 93% of intervention weeks	NR	No	Accelerometry
Losina et al, 2017 [39]	Flex	Daily for duration of intervention	Minimum of 10 h/day	NR	NR	N/A	Self-reporting
Mahar et al, 2015 [40]	Flex	Daily for duration of intervention	NR	NR	NR	No	Self-reporting
Measurement only							
Chen and Pu, 2014 [41]	Ultra and One	Daily for 2 weeks	NR	NR	Daily reminder to share experience of wearing Fitbit	No	NR
Pagkalos et al, 2017 [42]	Zip	Daily for duration of intervention	NR	NR	NR	No	Self-reporting
Ptomey et al, 2018 [43]	Charge HR	During intervention sessions	NR	100% (with staff supervision)	NR	No	NR

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Walsh and Golbeck, 2014 [44]	Classic	Daily for 10 days	NR	73% of participants were adherent	NR	Yes; same use	IPAQ
Yoon et al, 2018 [45]	Flex	Daily for duration of intervention	NR	Days participants were adherent: 66%	NR	Yes; same use	Self-reporting
Intervention and measurement							
Choi et al, 2016 [46]	Ultra	Daily for at least 10 h	Minimum of 1000 steps/day	Days participants were adherent: intervention: 78%; comparison: 80%	Participants entered steps into their daily activity diary	Yes; same use	Self-reporting
Chung et al, 2016 [47]	Zip	Daily for duration of intervention	NR	Days participants were adherent: overweight group: 99%; normal weight group: 78%	Study team sent Twitter message reminders	N/A	NR
Gilmore et al, 2017 [48]	Zip	Daily	NR	NR	NR	No	NR
Halliday et al, 2017 [49]	NR	Daily for duration of intervention	100 or more steps per day	50.5%-82.9% of participants adhered to wearing Fitbit on a weekly basis	Participants were invited to join a private group on the Fitbit website that allowed for data sharing	N/A	NR
Florence et al 2016 [50]	Flex	Daily for duration of intervention	NR	NR	NR	Yes; control group started Fitbit Flex on week 8	IPAQ
Miragall et al, 2017 [51]	One	Daily for duration of intervention	NR	N/A	N/A	Yes; blinded	NR
Schrager et al, 2017 [52]	Flex	Daily for duration of intervention	100 or more steps per day	Median number of eligible days where the participant recorded at least 100 steps was 27.5 (IQR 8)	Participants were given a 2-week acclimatization period to wear and use the device	N/A	Self-reporting of PA
Thorndike et al, 2014 [53]	Classic	Duration of intervention	500 or more steps/day	Percentage of worn days in each phase: 77% in phase 1 and 60% in phase 2	Weekly reminder emails to charge device and monetary incentives for high compliance rates	Yes; blinded	NR
Washington et al, 2014 [54]	Classic	Daily for duration of intervention	NR	2 subjects had missing Fitbit data	Participants earned opportunities to draw prizes and brought the device to the lab 3 times a week for charging and retrieving data	N/A	Self-reporting of PA
West et al, 2016 [55]	Zip and Aria	Daily for duration of intervention	NR	Students used their Fitbit for an average of 23.7 days (SD 15.2 days)	NR	No	NR

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Zhang and Jemmott, 2019 [56]	Zip	Daily for duration of intervention	NR	16% of Fitbit data were missing during intervention period	Daily notifications to wear Fitbit and log PA	Yes; same use	NR
Middle adulthood (41-64 years)							
Intervention only							
Amorim et al, 2019 [57]	NR	Daily	N/A	96% reported wearing every day or most days	NR	No	Accelerometry and IPAQ
Butryn et al, 2014 [58]	Flex	Daily for duration of intervention	NR	Participants wore 86% of days during intervention	Public display of PA data	N/A	GT3X+accelerometers
Cadmus-Bertram et al, 2015 [59]	One	Daily for duration of intervention	Minimum of 2000 steps/day	NR	NR	No	Accelerometry
Cadmus-Bertram et al, 2019 [60]	Charhe HR or Charge 2	Daily	N/A	NR	In-person instruction on Fitbit use	No	Accelerometry
Dean et al, 2018 [20]	Flex	Daily; duration of intervention	NR	Participants who were adherent to wear instructions: 70%	Participants received 3 text messages weekly	N/A	Community Health Activities Model Program for Seniors Questionnaire
Duncan et al, 2020 [61]	Alta	NR	NR	NR	NR	Yes, for both intervention groups; no, for control group	Accelerometry and Active Australia Survey
Ellingson et al, 2019 [62]	Charge	Use at participants' discretion for duration of intervention	Minimum of 10 h/day	NR	Intervention group determined cues to remember to wear Fitbit and check data	Yes; same use	Accelerometry
Kandula et al, 2017 [63]	Zip	Daily	NR	NR	NR	N/A	Actigraph Accelerometer and self-reported questionnaire
Ross and Wing, 2016 [64]	Zip and Aria	Daily	NR	Days participants were adherent: Tech: 76%; Tech+phone: 86%	Fitbit sent weekly emails updating progress	Fitbit used in one comparison group but not the other (pedometer used)	NR
Singh et al, 2020 [65]	Charge	As desired to self-monitor and manage PA	NR	Average h worn: 17.3 h (SD 5.7 h) per 6.1 days (SD 0.8 days) per week	Basic instruction on using and setting up Fitbit	No	Accelerometry and Active Australia Survey
Van Blarigan et al, 2019 [66]	Flex	Daily	NR	Participants wore Fitbit for 88% of study days	N/A	No	Accelerometry
Measurement only							

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Patel et al, 2017 [67]	Flex	Daily	At least 1000 steps/day	10.1% of missing observation days in intervention arm and 12.7% in control arm	NR	Yes	NR
Robinson et al, 2019 [68]	Zip	Daily during waking hours	NR	NR	Participants asked to sync Fitbit data daily	Yes; same use	NR
Schumacher et al, 2017 [69]	Flex	Daily	Minimum of 100 steps/day	97% adherent to wear time criteria	NR	N/A	NR
Intervention and measurement							
Adams et al, 2017 [70]	Zip	Daily during waking hours	NR	NR	Text step counts daily and random selection for monthly incentives for wearing their Fitbit regularly	Yes	IPAQ
Arigo, 2015 [71]	Flex	Daily; duration of intervention	NR	Days participants were adherent: 93%	Badges for achieving PA milestones; participants were advised to check step progress daily	N/A	NR
Arigo et al, 2015b [72]	Flex	Daily for duration of intervention	Defined as >100 steps in a day	Participants wore 97% of days during intervention	Instructions on device use, public display of steps data, and PA partner accountability	NA	NR
Finkelstein et al, 2015 [73]	One	Daily	NR	3 participants did not provide Fitbit data	Instructions and use of device before study for comfort and familiarity	Yes	Self-reporting
Fukuoka et al, 2018 [74]	Zip	Daily	Minimum of 8 h/day	NR	NR	N/A	IPAQ short version
Gell et al, 2020 [75]	One	Daily for duration of intervention	Minimum of 10 h/day	Average days participants were adherent: 6 days/week	NR	Yes; same use	Accelerometry
Gremaud et al, 2018 [76]	Zip	Daily during waking hours	NR	64.6% wear time in Fitbit arm with a 16.5% increase for Fitbit+Map Trek arm	Reminder system, which prompted each user to wear their Fitbit following nonwear days	Yes	NR
Grossman, et al 2017 [77]	Charge HR	Duration of intervention	NR	NR	NR	Yes	NR
Linke et al, 2019 [78]	Charge HR	Daily for duration of intervention	NR	NR	Participants met with study team to sync Fitbit weekly and problem-solve Fitbit-related issues	N/A	Godin Leisure-Time Exercise Questionnaire
Meints et al, 2019 [79]	Flex	Duration of intervention	Minimum of 10 h/day and 4 days/week	18 (out of 26) average valid weeks of Fitbit wear	Participants earned monetary reward for accurate use of Fitbit during first 2 weeks	N/A	NR

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Painter et al, 2017 [80]	NR	Daily use	NR	NR	NR	NR	NR
Reed et al, 2019 [81]	Charge 2	Daily during waking hours	NR	NR	Basic instruction on using and setting up Fitbit	Yes; same use	Godin Leisure-Time Exercise Questionnaire
Wang et al, 2015 [82]	One	Duration of intervention	Minimum of 10 h/day	Nontypical days (not meeting wear time criteria) ranged from 5%-9%	NR	Yes	Accelerometry
Willis et al, 2017 [83]	Flex	Daily	NR	NR	NR	Yes	Accelerometry and self-reporting
Older adulthood (≥65 years)							
Intervention only							
Ashe et al, 2015 [84]	One	Daily for 26 weeks	NR	NR	NR	No	Accelerometry
Christiansen et al, 2020 [85]	Zip	Daily during waking hours	NR	60% of intervention group monitored steps at least 80% of study time	In-person instruction of Fitbit use	No	Accelerometry
Kenfield et al, 2019 [86]	One	Duration of intervention	NR	Fitbits worn 98% of days during intervention	NR	No	Accelerometry and self-reporting
Thompson et al, 2014 [21]	NR	Daily for 48 weeks	NR	NR	NR	Yes; same use	Accelerometry
Measurement only							
Rossi et al, 2018 [87]	Alta	At all times for 30 days; remove only for bathing and sleeping	NR	Participants wore median of 93% of 30 days	Staff called participants after 1 week	N/A	Godin Leisure-Time Exercise Questionnaire
Schmidt et al, 2018 [88]	Charge HR	14 consecutive days during waking hours	NR	2 participants excluded for not wearing the device for a week	3 home visits	N/A	NR
Streber et al, 2017 [89]	Zip	During waking hours for 7 consecutive days	Minimum of 8 h/day	NR	No charging and no turning off and on	Yes; same use	Self-reporting
Intervention and measurement							
Harkins et al, 2017 [90]	Ultra	Daily	NR	NR	Daily email or text message and financial incentives for meeting goal	Yes; same use	Self-reporting
McMahon et al, 2017 [91]	One	During waking hours for 7 consecutive days	NR	Average hours worn at baseline: 13.01 (SD 1.87)	Participants asked to document days or times monitor was used; staff reviewed documentation and data	Yes; same use	Community Health Activities Model Program for Seniors Questionnaire

Study	Fitbit	Wear instructions	Fitbit use adherence			Fitbit used in comparison group?	Other PA[a] measures
			Minimum wear time criteria	Rate	Strategies to boost adherence		
Vidoni et al, 2016 [92]	Zip	During waking hours	NR	NR	Staff made biweekly phone calls and additional calls if no activity for 3 days	Yes; device masked for 8 weeks versus 1 week	6-min walk test, mini-physical performance test, and battery of timed physical tasks

[a]PA: physical activity.

[b]Only the reported intervention condition data.

[c]NR: not reported.

[d]N/A: not applicable.

[e]IPAQ: International Physical Activity Questionnaire.

Childhood (9-12 Years)

General Study Characteristics

The 6 childhood studies had sample sizes ranging from 6 to 116 participants and were either single-group (n=3) or quasi-experimental designs (n=3). All studies were conducted in a school setting, and when appropriate, tried to integrate the intervention sessions into regular, daily school activities, including class sessions and recess periods. The most commonly used behavior change techniques were goal setting (through individual and group challenges) and positive reinforcement (through rewards). The duration of the intervention ranged between 4 and 12 weeks.

Fitbit Use

The most commonly used Fitbit model was the Fitbit Charge, which was used in 4 of the 6 interventions [18,23,25,26]. A total of 3 studies used Fitbits for both intervention and measurement purposes, 2 for intervention only, and 1 for measurement only. Participants in the comparison condition used Fitbit devices in only one of the 3 quasi-experimental studies.

Wear Time and Adherence

In total, 5 of the 6 interventions instructed participants to wear the device for a specific period. A total of 2 studies restricted device wear time to in-school supervised intervention sessions and reported that 100% of participants adhered to the device wear protocol, largely because of study staff monitoring [24,27]. The 2 interventions instructed participants to wear their Fitbits only during school days for the duration of the intervention [18,26]. In one study, participants were asked to wear the device for 24 hours during a 7-day period [25]. Applying a wear time criterion of 8 hours per day, one study reported that participants were adherent on 65%-73% of intervention days [23].

Adolescence (13-17 Years)

General Study Characteristics

The 11 adolescent studies had sample sizes ranging from 6 to 388 participants. In total, 6 of the interventions used a randomized controlled trial design, 3 were quasi-experimental, and 2 used a single-group design. In total, 4 studies used an electronic or web-based platform for intervention delivery, including 3 that used mobile apps for data collection and the delivery of intervention content [28,29,33,36] and 1 that used Facebook as a web-based platform to encourage interactions between participants [30]. A total of 7 studies were delivered in a school setting [19,25,31-34,36]. Across all studies, the most commonly used behavioral change techniques were goal setting, self-monitoring, and knowledge shaping. The study duration varied between 4 weeks and 24 months.

Fitbit Use

The most commonly used Fitbit model was the Fitbit Flex, which was used in 5 of the 12 interventions [28,30,33,34,36]. The Fitbit Zip was the second most commonly used device (in 3 studies [31,32,35]). A total of 5 studies used Fitbits for both intervention and measurement purposes, 3 for intervention only, and 3 for measurement only. In 7 of the 10 studies with multiple conditions, participants in the comparison condition used Fitbit devices.

Wear Time and Adherence

Overall, 5 studies instructed participants to wear the device daily for the entire duration of the study [19,28,30,33,34], 4 studies instructed participants to wear the device for 7-day data collection periods only [25,32,35,36], and the remaining 2 studies did not report wear instructions [29,31,48]. Moreover, 5 studies used a minimum wear time criterion that was defined by either the number of hours (eg, 8 hours, 10 hours, or 24 hours per day) or steps (eg, 500 or 1000 steps per day) [19,30,32,35,36]. In addition, 3 studies reported the percentage of intervention days on which a specific minimum wear criterion was met (67.3% [19], 71.5% [30], and 73.4% [36]). One study excluded participants from the analysis who did not meet the wear time criterion [32]. One intervention that did not use the minimum wear time criterion was able to report an average number of days of valid Fitbit wear of 78.1 (SD 8.6; of a maximum of 84 days) for intervention participants [34]. Another study without a minimum wear time criterion reported that 36% of participants never wore their Fitbit [33].

Strategies to Boost Wear Time

Strategies to boost wear time included providing participants with oral and written instructions for Fitbit use [19,32]. Some studies also sent participants daily or weekly text messages or emails to encourage consistent use, meeting PA goals, or data upload [28-30,32]. In one study, a weekly lottery was used to reward participants with gift cards [33].

Other Measures of PA

Furthermore, 3 studies assessed PA with accelerometers at data collection time points [19,30,34], and 3 studies used self-report measures of PA [19,25,28].

Early Adulthood (18-40 Years)

General Study Characteristics

The 20 eligible studies for adults aged 18-40 years had a range of sample sizes of participants. Randomized controlled trials (RCTs) were the most commonly used study design (11/20, 55% studies), followed by single-group study designs (5/20, 25% studies). In total, 12 of the 20 studies used mobile apps, web-based platforms, emails, or text messages for intervention delivery [41-44,46-51,55,56]. Of these studies, 3 encouraged web-based interactions between participants [41,44,47]. In total, 8 of the 20 studies used a campus- or workplace-based approach to intervention delivery [37,39,40,50-53,55]. Strategies for behavioral change included competition or challenges, both at the individual and group levels, and self-monitoring, social support, and goal setting. The study duration ranged from 1 week to 12 months.

Fitbit Use

The most commonly used Fitbit models were Fitbit Zip and Flex, which were used in 11 of the 20 studies [37,39,40,42,45,47,48,50,52,55,56]. Furthermore, 10 studies used Fitbits for both intervention and measurement purposes, 4 for intervention only, and 5 for measurement only. In 6 of the 15 studies with multiple conditions, participants in the comparison condition used Fitbit devices.

Wear Time and Adherence

All but 3 studies [37,41,44] instructed participants to wear the device daily, either at all times or during waking hours, for the duration of the intervention. Furthermore, 2 studies instructed participants to wear the device for a specific data collection period [41,44]. Different metrics were used to report adherence to daily wear instructions. There were 3 studies that reported the percentage of intervention days in which participants were adherent: 66% [45], 73% [44], and 78%-99% [47]. Another study reported that, on average, participants were adherent on 23.7 (SD 15.2) days (of 63 days) [55]. One study instructed participants to wear the device only during intervention sessions, and 100% of the participants were adherent [43]. Minimum wear time criteria were also used to report adherence. One study with a minimum wear time criterion of 1000 steps per day reported that participants met the criterion on 78% of intervention days [46], whereas another study in which the minimum wear time criterion was set at 500 steps per day reported that participants met the criterion on 60%-70% of intervention days [53]. A minimum criterion of 100 steps per day allowed one study to report a median number of 27.5 days (of 30) on which participants were adherent [52]. Another study with the same minimum wear time criterion reported that 51%-83% of participants were adherent [49]. With a minimum wear criterion of one day per week, one study reported that participants were adherent on 93% of intervention weeks on average [38].

Strategies to Boost Wear Time

Strategies to boost wear time included sending daily emails to inquire about Fitbit use experience [41], prompting participants to enter daily Fitbit data into an app [46], asking participants to share Fitbit data publicly [49], or sending daily reminder messages and instructions on Fitbit use [47]. Some studies provided participants with opportunities to win incentives based on compliance rates [53,54].

Other Measures of PA

A total of 10 studies asked participants to self-report their PA using instruments such as the International PA Questionnaire, the Stanford Brief PA Survey, and the 30-day PA Recall [37,39,40,42,44-46,50-52,54-56]. Only 1 study used an additional objective measure of PA (ie, accelerometer [38]).

Middle Adulthood (41-64 Years)

General Study Characteristics

The sample sizes in the 28 middle adulthood studies ranged from 11 to 2113 participants. Most of the studies were RCTs (17/28, 61%), and 20 interventions used technology (eg, texts, apps, and social media) for intervention delivery [57,58,61,63,64,66,67,70-77,80-83,93]. The most common behavior change techniques used were self-monitoring, social support, behavioral counseling, and goal setting. The study duration ranged from 4 weeks to 6 months.

Fitbit Use

The most commonly used device was the Fitbit Flex, which was used in 9 studies [20,58,66,67,69,71,72,79,83]. There were 14 studies that used Fitbit for both intervention and measurement purposes, 11 for intervention only, and 3 for measurement only. Of the 18 studies with multiple conditions, 13 provided participants in the comparison condition with Fitbit devices.

Wear Time and Adherence

All but 3 studies [61,62,65] instructed participants to wear the device daily, either at all times or during waking hours, for the duration of the intervention. Among them, 2 studies reported the percentage of participants who were adherent to daily wear instructions: 96% [57] and 70% [20]. Other studies reported the percentage of days on which participants were adherent to wear instructions: 86% [58], 88% [66], 97% [71], 93% [72], and 76%-86% [64]. Furthermore, 9 studies also used a minimum wear time criterion defined by either the number of hours (eg, 8 or 10 hours per day) or steps (eg, 100 or 2000 steps per day) [59,62,67,69,71,74,75,79,82]. With a minimum wear time criterion of 100 steps per day, 1 study reported that 97% of the participants were adherent [69]. A minimum wear criterion of 10 hours per day allowed another study to report 18 of 26 average valid weeks of Fitbit wear [79], whereas another study

used the same criterion to report that participants were adherent to the criterion on 6 days per week on average [75]. A minimum criterion of 10 hours per day was also used in another study to report 5%-9% of days on which participants did not meet the criterion on average [82]. Similarly, with a minimum wear time criterion of 1000 steps per day, another study reported 10.1%-12.7% of missing observation days [67]. Allowing participants to self-monitor PA as desired, one study reported the average hours worn of 17.3 (SD 5.7) hours per 6.1 (SD 0.8) days per week [65]. Another study excluded 3 participants who provided no Fitbit data [73].

Strategies to Boost Wear Time

Various strategies were used to promote Fitbit wear, including weekly texts to encourage PA based on Fitbit data [20], weekly emails providing activities' progress summaries [64], asking participants to sync Fitbit data daily [68], providing incentives for wearing Fitbit regularly [70], public display of Fitbit data [58,71], and instructions on device use [71,73].

Other Measures of PA

Objective measures to assess PA were used in 12 studies [57-63,65,66,75,82,83], whereas self-reported measures were used in 11 studies [20,57,61,63,65,70,73,74,78,81,83].

Older Adulthood

General Study Characteristics

The 10 older adulthood studies had sample sizes ranging from 25 to 102 participants, and most (8/10, 80%) were RCTs. Studies with older adults used individual and group-based approaches for intervention delivery. In addition to encouraging individualized PA goal setting or prescribing exercises, 3 studies involved regular phone calls made by study counselors or coaches [21,85,92]. One study provided participants with access to a study website and used text messages for intervention delivery [86]. Interventions providing PA education were often delivered in a group setting through a community-based approach, which allowed for the use of social support as a behavioral change technique [84,89,91]. Other behavioral change techniques included goal setting, behavioral counseling, and self-monitoring.

Fitbit Use

Different Fitbit devices were used across studies, including Classic, Zip, Ultra, Charge HR, and One, with none being predominant. In addition, 3 studies used Fitbit for both intervention and measurement purposes, 4 for intervention only, and 3 for measurement only. Of the 8 studies with multiple conditions, 5 provided participants in the comparison condition with Fitbit devices.

Wear Time and Adherence

All but 2 studies [89,91] instructed participants to wear the device daily, either at all times or during waking hours, for the duration of the intervention. Using daily wear instructions, the number of days the device worn was commonly reported either as an average (6.6, SD 1.1 over 7 days) [91] or as a median (93% over 30 days) [94]. One study reported that 60% of participants in the intervention group used Fitbit at least 80% of the study time [85], whereas another study simply reported that Fitbit was worn on 98% of days during the intervention [86]. One study used a minimum wear time criterion (8 hours per day) but did not report adherence to the criterion [89]. One study excluded 2 participants who did not wear the device for at least half of the instructed wear period (14 days) [88].

Strategies to Boost Wear Time

Strategies used to promote wear time adherence included providing participants with wear instructions and reminders via phone calls and text messages [85,87,90,92]. Some studies also asked participants to upload PA data on a daily basis or to document the device wear time and day [89,91].

Other Measures of PA

All but one study [88] used an additional measure of PA. Although self-reporting (using different scales) was the most common measure, which was used in 6 studies [86,89-92,94], accelerometers were used in 4 studies [21,84-86]. One study used a physical performance test along with a walk test [92].

Discussion

Principal Findings

This study reviewed the use of Fitbit devices in PA intervention studies across the life course. In addition to differences in study designs and intervention delivery methods, our results indicate considerable heterogeneity in Fitbit use within and between developmental stages. From early to older adulthood, most studies instructed participants to wear their Fitbit daily, either at all times or during waking hours, for the duration of the intervention. Studies conducted among children and adolescents tended to specify more limited device wear periods (eg, 24 hours for 7 days). Within developmental stages, our findings also suggest a lack of consistency in the definition of wear time criteria, which sometimes were used to report different adherence metrics or to exclude incomplete data from study analyses. A total of 8 different types of Fitbit devices were used across all age groups, with Fitbit Flex and Zip being the most predominant and some seemingly discontinuing use as newer devices became available. Regardless of intended Fitbit use (ie, measurement vs intervention tool), strategies to boost wear time were similar across stages, and the most commonly used strategies included sending participants reminders through texts or emails and asking participants to log their steps or sync their Fitbit data daily. Overall, the heterogeneity in Fitbit use across PA intervention studies reflects its relative novelty in the field of research.

Across all stages, based on the taxonomy developed by Lyons et al [95], the most common behavior change techniques used were self-monitoring and goal setting, regardless of the intended device use. This aligns with previous findings indicating goal setting and self-monitoring as the most commonly used behavior change techniques in studies with activity trackers [96]. As a self-monitoring technology, Fitbit devices provide real-time feedback that has the potential to stimulate behavior change. Self-monitoring allows participants to establish and track goals that were commonly operationalized through individual or group step count challenges. For example, a classroom-based study

in children used individual step goals consistent with achieving 60 minutes of moderate-to-vigorous physical activity (MVPA) per day [23]. Additional behavioral change techniques appeared to be developmentally targeted. For example, among children, rewards for meeting step goals were often provided (eg, accruing points toward gift card balance). Through the use of social media platforms, adolescents and adults were provided with performance-based, web-based badges [41,97]. Among older adults, group-based PA education along with individual PA coaching or counseling provided social support to encourage the initiation and maintenance of behavior change [89].

Similar to behavior change techniques, the heterogeneity we observed regarding wear instructions and criteria also seemed to be because of developmental considerations. Most studies conducted among children and adolescents opted for instructions that required the device to be worn daily (8-24 hours) for a set data collection period (5-14 days); these studies did not set specific wear time criteria for inclusion in the analyses. Our findings align with previous results indicating a considerable reduction in the use of wearable trackers in youth following the first 2 weeks [19,98]. As such, limited device wear time in children and adolescents could potentially be a strategy that aims at capitalizing on wear patterns and usability trends in these groups. Studies conducted during early and middle adulthood tended to specify a minimum wear time criterion for inclusion in analyses based on specific numbers of steps or hours, in addition to daily wear instructions. However, studies conducted in older adults did not set minimum wear time criteria and instructed participants to wear the device daily during waking hours. The less rigid guidelines for device wear adherence among older adults could potentially be a way of increasing feasibility in populations who are less able to meet strict criteria and are less proficient in the use of technology [99].

Despite the importance of meeting a minimum threshold of wear time criteria to calculate a reliable estimate of PA, the results from this review also indicated a lack of consistency in the criteria used to define adherence to device wear *within* developmental stages. A systematic review that examined the length of device wear time required in PA interventions found that most studies conducted among adults did not report minimum device wear and that there was significant variation among studies reporting these criteria [22]. Corresponding to the lack of uniformity in wear time criteria, different metrics (eg, percentage, mean, and median) were used to report rates of adherence to wear instructions. If not met, the wear time criterion was sometimes used to exclude participants from the data analysis. However, many studies used the wear time criteria to report different metrics of adherence. Overall, the absence of clear reporting with standardized metrics significantly impaired efforts to assess overall adherence rates within developmental stages.

The most common pattern that emerged across studies was the use of reminder strategies to boost wear time, which did not differ by the intended device use (ie, intervention or measurement). Generally, texts and emails were sent on a daily or weekly basis as PA and Fitbit wear reminders. Manually logging or syncing Fitbit data on a daily basis was also a strategy to indirectly promote Fitbit wear on a daily basis. Results from previous studies indicate that, in addition to forgetting to wear their trackers [100], approximately 2% of study participants stopped using their devices each week altogether [101], and study participants also reported using their Fitbit less than 10% of the time following the end of wear-based incentives [102]. Therefore, these strategies are particularly essential given the evidence regarding decrease in Fitbit wear adherence over time in users and the need for reminder strategies to boost wear time [103].

Despite questions regarding the validity of Fitbits for assessing PA [104], most interventions in this review used Fitbit devices for both intervention and measurement purposes (39/75, 52%) or for data measurement purposes exclusively (15/75, 20%). Most studies (45/75, 63%) that were reviewed supplemented the use of Fitbit with additional objective (eg, accelerometers) or self-reported (eg, International PA Questionnaire) measures of PA. It is possible that the addition of other PA measures, even in studies that used Fitbit devices primarily as a measurement or data collection tool, was because of concerns about the uncertainty around the accuracy of measures provided by Fitbit devices [104]. In addition, the use of other measures (ie, accelerometry or self-reporting) to collect baseline or habitual activity [48] could also point to the perceived inaccuracy of data collected from commercially available trackers, which could have a potential impact on activity. Previous studies have also shown that commercially available trackers such as Fitbit devices often overestimate the time spent in MVPA compared with research-grade monitors [15,104,105].

However, the use of additional PA measures is not limited to addressing the accuracy issues. Results from a recent systematic review and meta-analysis of Fitbit-based interventions highlighted that the use of accelerometers and self-report, in addition to Fitbit, is often done to capture PA outcomes other than steps [106]. With the expansion of the use of Fitbit devices in PA intervention studies, previous studies have raised issues regarding their inability to capture PA constructs such as nonambulatory activities or energy expenditure [107]. In a recently published paper, Balbim et al [108] summarized the challenges and possible solutions to use Fitbit devices in mobile health intervention research. They described challenges and solutions at four different study phases: preparation, intervention delivery, data collection and analysis, and study closeout. For example, during the data collection phase, they point to the inaccuracy or unavailability of wear time data through Fitbit's web API. They then discussed the potential solution of using heart rate data and pre-established rules for determining wear time and manually identifying gaps in heart rate data, indicating nonwear time. They also highlight the tedious and challenging nature of such an endeavor [108]. Thus, the use of additional PA measures (objective and subjective), despite increased burden on participants, allows for the efficient collection of different types of data, including valid wear time, information about body positions, sedentary behaviors, postural allocation, and the type of activity being performed [107,109-111].

Strengths and Limitations

The primary limitation of this review is that the search for articles was restricted to articles available in the Fitabase library between 2012 and 2018 or on PubMed between 2019 and 2020. Given that the Fitabase library uses the systematic searching procedures of several databases (eg, PubMed, Google Scholar, and Science Direct), searching only PubMed for articles from 2019 to 2020 could have resulted in missed literature. In addition, this review was limited to intervention studies published in English and likely missed formative work that could provide important information regarding the design of Fitbit-based studies. Despite these limitations, this review provides insight into the current state of affairs in Fitbit use in research by focusing on different developmental stages and how the use of the device differs across those stages. Describing both study characteristics and the use of Fitbit devices provides insight into PA study designs across the lifespan and the different ways in which these monitoring devices are used.

Conclusions

Insufficient PA across the lifespan is associated with an increased risk of numerous chronic diseases and is a major public health issue [1]. The prominence and relatively low cost of Fitbit devices have increased their use by the public and researchers as PA trackers. Although behavior change techniques and strategies to boost Fitbit wear time were similar across all studies reviewed, our findings indicate significant differences in wear instructions and metrics for reporting adherence rates. Although between-stage differences appear to be based on developmental considerations that aim to maximize device use in each age group, within-group differences appear to result from a lack of uniformity in metrics used to report rates of adherence and minimum wear time criteria. The use of additional PA data collection tools in most studies that were reviewed points to the accuracy issues raised by previous research focusing on Fitbits in PA interventions [104,105] and a reluctance to rely on Fitbits as the primary measurement device or for the assessment of habitual activity. However, additional PA measures are also used to capture PA constructs not measured by Fitbit devices (eg, MVPA, sedentary behaviors, and types of activity). As the use of monitoring devices continues to expand in the field of PA research, the lack of uniformity in study protocols and metrics of reported measures represents a major issue for purposes of comparison [112]. Given that clinical trial registries serve as a repository for researchers [113], there is a need for increased transparency in the prospective registration of PA intervention studies. This paper serves as a call for researchers using Fitbit devices to provide a clear rationale for the use of several PA measures and to specify the metrics that will be reported for each. By providing researchers with a synthesis of information on the use of Fitbit devices in PA intervention studies across the life course, this narrative review serves as a resource that may be used to inform the design of future trials involving Fitbit devices.

Conflicts of Interest

None declared.

References

1. Jake-Schoffman D, Baskin M. Promoting physical activity across the life span: progress and promise. J Public Health Manag Pract 2019;25(1):30-31. [doi: 10.1097/PHH.0000000000000927] [Medline: 30507802]
2. Guthold R, Stevens G, Riley L, Bull F. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):1077-1086. [doi: 10.1016/S2214-109X(18)30357-7]
3. Booth F, Roberts C, Laye M. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211 [FREE Full text] [doi: 10.1002/cphy.c110025] [Medline: 23798298]
4. Chen Y, Sloan FA, Yashkin AP. Adherence to diabetes guidelines for screening, physical activity and medication and onset of complications and death. J Diabetes Complicat 2015 Nov;29(8):1228-1233. [doi: 10.1016/j.jdiacomp.2015.07.005]
5. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
6. Worldwide quaterly wearable device tracker. IDC Corporate. URL: https://www.idc.com/tracker/showproductinfo.jsp?containerId=IDC_P31315 [accessed 2021-05-04]
7. Fitbit. URL: https://www.fitbit.com/us/home [accessed 2019-12-09]
8. Vailshery LS. Fitbit unit sales worldwide 2010-2019. Statista. 2021. URL: https://www.statista.com/statistics/472591/fitbit-devices-sold/ [accessed 2021-05-05]
9. Fitbit reports 2019 fourth quarter and full year results. Fitbit, Inc. 2020. URL: https://investor.fitbit.com/press-releases/press-release-details/2020/Fitbit-Reports-2019-Fourth-Quarter-and-Full-Year-Results/default.aspx [accessed 2021-05-05]
10. Brinton JE, Keating MD, Ortiz AM, Evenson KR, Furberg RD. Establishing linkages between distributed survey responses and consumer wearable device datasets: a pilot protocol. JMIR Res Protoc 2017 Apr 27;6(4):e66 [FREE Full text] [doi: 10.2196/resprot.6513] [Medline: 28450274]
11. Fitabase. URL: https://www.fitabase.com/ [accessed 2019-12-10]
12. Guo F, Li Y, Kankanhalli M, Brown M. An evaluation of wearable activity monitoring devices. In: Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia. 2013 Presented at: MM '13: ACM Multimedia Conference; October, 2013; Barcelona Spain p. 31-34. [doi: 10.1145/2509352.2512882]

13. Dannecker K, Sazonova N, Melanson E, Sazonov E, Browning R. A comparison of energy expenditure estimation of several physical activity monitors. Med Sci Sports Exerc 2013 Nov;45(11):2105-2112 [FREE Full text] [doi: 10.1249/MSS.0b013e318299d2eb] [Medline: 23669877]
14. McArdle WD, Katch FI, Katch VL. Accuracy of the Fitbit pedometer for self-pacedprescribed physical activity. In: Ramirez E, Peterson C, Wu W, Norman G, editors. Sports and Exercise Nutrition 4th Edition. Philadelphia, Pennsylvania, United States: Lippincott Williams & Wilkins; 2012:1-681.
15. 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 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
16. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
17. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport 2014 Sep;17(5):496-500. [doi: 10.1016/j.jsams.2013.10.241] [Medline: 24268570]
18. Buchele Harris H, Chen W. Technology-enhanced classroom activity breaks impacting children's physical activity and fitness. J Clin Med 2018 Jun 29;7(7):165 [FREE Full text] [doi: 10.3390/jcm7070165] [Medline: 29966308]
19. Gaudet J, Gallant F, Bélanger M. A Bit of Fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
20. Dean DA, Griffith DM, McKissic SA, Cornish EK, Johnson-Lawrence V. Men on the move-nashville: feasibility and acceptability of a technology-enhanced physical activity pilot intervention for overweight and obese middle and older age African American men. Am J Mens Health 2018 Jul 19;12(4):798-811 [FREE Full text] [doi: 10.1177/1557988316644174] [Medline: 27099346]
21. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014 May;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]
22. Jake-Schoffman DE, Silfee VJ, Sreedhara M, Rosal MC, May CN, Lopez-Cepero A, et al. Reporting of physical activity device measurement and analysis protocols in lifestyle interventions. Am J Lifestyle Med 2019 Jul 17:155982761986217. [doi: 10.1177/1559827619862179]
23. Evans E, Abrantes A, Chen E, Jelalian E. Using novel technology within a school-based setting to increase physical activity: a pilot study in school-age children from a low-income, urban community. Biomed Res Int 2017;2017(4271483) [FREE Full text] [doi: 10.1155/2017/4271483] [Medline: 29670894]
24. Mackintosh K. Mission possible: using ubiquitous social goal sharing technology to promote physical activity in children. Malaysian J Move, Health Exerc 2016 Oct 18;5(2). [doi: 10.15282/mohe.v5i2.115]
25. Walther A, Dunker T, Franzen-Castle L, Krehbiel M. Increasing at-risk youth self-reported and objectively measured physical activity in an afterschool program. J Fam Consum Sci 2018 Mar 01;110(1):59-63. [doi: 10.14307/jfcs110.1.59]
26. Harris HB, Cortina KS, Templin T, Colabianchi N, Chen W. Impact of coordinated-bilateral physical activities on attention and concentration in school-aged children. Biomed Res Int 2018 May 28;2018:2539748 [FREE Full text] [doi: 10.1155/2018/2539748] [Medline: 29998131]
27. Hayes LB, Van Camp CM. Increasing physical activity of children during school recess. J Appl Behav Anal 2015 Sep 29;48(3):690-695. [doi: 10.1002/jaba.222] [Medline: 26119136]
28. Chen J, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017 Aug 02;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
29. Gandrud L, Altan A, Buzinec P, Hemphill J, Chatterton J, Kelley T, et al. Intensive remote monitoring versus conventional care in type 1 diabetes: a randomized controlled trial. Pediatr Diabetes 2018 Feb 21;19(6):1086-1093. [doi: 10.1111/pedi.12654] [Medline: 29464831]
30. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec 15;64(12):e26660. [doi: 10.1002/pbc.26660] [Medline: 28618158]
31. Haegele J, Porretta D. A leisure-time physical activity intervention for adolescents with visual impairments. Res Q Exerc Sport. 2016. URL: https://search.proquest.com/openview/e1e94139646d164224809faee1a60f4c/1?pq-origsite=gscholar%7B%5C&%7Dcbl=40785 [accessed 2021-05-07]
32. Meng Y, Manore M, Schuna J, Patton-Lopez M, Branscum A, Wong S. Promoting healthy diet, physical activity, and life-skills in high school athletes: results from the WAVE ripples for change childhood obesity prevention two-year intervention. Nutrients 2018 Jul 23;10(7):947 [FREE Full text] [doi: 10.3390/nu10070947] [Medline: 30041446]
33. Pope L, Garnett B, Dibble M. Lessons learned through the implementation of an eHealth physical activity gaming intervention with high school youth. Games Health J 2018 Apr;7(2):136-142. [doi: 10.1089/g4h.2017.0164] [Medline: 29393679]

34. Remmert JE, Woodworth A, Chau L, Schumacher LM, Butryn ML, Schneider M. Pilot trial of an acceptance-based behavioral intervention to promote physical activity among adolescents. J Sch Nurs 2019 Dec 13;35(6):449-461 [FREE Full text] [doi: 10.1177/1059840518786782] [Medline: 30004269]
35. Short KR, Chadwick JQ, Cannady TK, Branam DE, Wharton DF, Tullier MA, et al. Using financial incentives to promote physical activity in American Indian adolescents: a randomized controlled trial. PLoS ONE 2018 Jun 1;13(6):e0198390. [doi: 10.1371/journal.pone.0198390]
36. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. A randomized controlled trial testing a social network intervention to promote physical activity among adolescents. BMC Public Health 2018 Apr 23;18(1):542 [FREE Full text] [doi: 10.1186/s12889-018-5451-4] [Medline: 29685112]
37. Bang K, Lee I, Kim S, Lim CS, Joh H, Park B, et al. The effects of a campus forest-walking program on undergraduate and graduate students' physical and psychological health. Int J Environ Res Public Health 2017 Jul 05;14(7):728 [FREE Full text] [doi: 10.3390/ijerph14070728] [Medline: 28678203]
38. Baruth M, Schlaff RA, Deere S, Walker JL, Dressler BL, Wagner SF, et al. The feasibility and efficacy of a behavioral intervention to promote appropriate gestational weight gain. Matern Child Health J 2019 Dec 20;23(12):1604-1612. [doi: 10.1007/s10995-019-02812-6] [Medline: 31541375]
39. Losina E, Smith SR, Usiskin IM, Klara KM, Michl GL, Deshpande BR, et al. Implementation of a workplace intervention using financial rewards to promote adherence to physical activity guidelines: a feasibility study. BMC Public Health 2017 Dec 01;17(1):921 [FREE Full text] [doi: 10.1186/s12889-017-4931-2] [Medline: 29195494]
40. Mahar M, Nanney L, Das B, Raedeke T, Vick G, Rowe D. Effects of an intervention using movement technology in a University physical activity class. Med Sci Sports Exerc 2015:522. [doi: 10.1249/01.mss.0000477869.71815.de]
41. Chen Y, Pu P. HealthyTogether: exploring social incentives for mobile fitness applications. In: Proceedings of the Second International Symposium of Chinese CHI. 2014 Presented at: Chinese CHI '14: The Second International Symposium of Chinese CHI; April, 2014; Toronto Ontario Canada p. 25-34. [doi: 10.1145/2592235.2592240]
42. Pagkalos I, Kokkinopoulou A, Weal M, Petrou L, Hassapidou M. Exercise monitoring of young adults using a Facebook application. Digit Health 2017 May 22;3:2055207617711286 [FREE Full text] [doi: 10.1177/2055207617711286] [Medline: 29942601]
43. Ptomey LT, Szabo AN, Willis EA, Gorczyca AM, Greene JL, Danon JC, et al. Changes in cognitive function after a 12-week exercise intervention in adults with Down syndrome. Disabil Health J 2018 Jul;11(3):486-490 [FREE Full text] [doi: 10.1016/j.dhjo.2018.02.003] [Medline: 29501470]
44. Walsh G, Golbeck J. StepCity: a preliminary investigation of a personal informatics-based social game on behavior change. In: Proceedings of the CHI '14 Extended Abstracts on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April, 2014; Toronto Ontario Canada p. 2371-2376. [doi: 10.1145/2559206.2581326]
45. Yoon S, Schwartz JE, Burg MM, Kronish IM, Alcantara C, Julian J, et al. Using behavioral analytics to increase exercise: a randomized N-of-1 study. Am J Prev Med 2018 Apr;54(4):559-567 [FREE Full text] [doi: 10.1016/j.amepre.2017.12.011] [Medline: 29429607]
46. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHhealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May 9;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
47. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to health: a novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2017 Jan 19;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
48. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized mobile health intervention for health and weight loss in postpartum women receiving women, infants, and children benefit: a randomized controlled pilot study. J Womens Health (Larchmt) 2017 Jul;26(7):719-727 [FREE Full text] [doi: 10.1089/jwh.2016.5947] [Medline: 28338403]
49. Halliday GC, Miles GC, Marsh JA, Kotecha RS, Alessandri AJ. Regular exercise improves the well-being of parents of children with cancer. Pediatr Blood Cancer 2017 Dec 19;64(12):e26668. [doi: 10.1002/pbc.26668] [Medline: 28627013]
50. H-Jennings F, Clément M, Brown M, Leong B, Shen L, Dong C. Promote students' healthy behavior through sensor and game: a randomized controlled trial. Med Sci Educ 2016 May 3;26(3):349-355. [doi: 10.1007/s40670-016-0253-8]
51. Miragall M, Domínguez-Rodríguez A, Navarro J, Cebolla A, Baños RM. Increasing physical activity through an Internet-based motivational intervention supported by pedometers in a sample of sedentary students: a randomised controlled trial. Psychol Health 2018 Apr 07;33(4):465-482. [doi: 10.1080/08870446.2017.1368511] [Medline: 28880576]
52. Schrager JD, Shayne P, Wolf S, Das S, Patzer RE, White M, et al. Assessing the influence of a Fitbit physical activity monitor on the exercise practices of emergency medicine residents: a pilot study. JMIR Mhealth Uhealth 2017 Jan 31;5(1):e2 [FREE Full text] [doi: 10.2196/mhealth.6239] [Medline: 28143805]
53. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. PLoS One 2014 Jun 20;9(6):e100251 [FREE Full text] [doi: 10.1371/journal.pone.0100251] [Medline: 24950218]

54. Washington WD, Banna KM, Gibson AL. Preliminary efficacy of prize-based contingency management to increase activity levels in healthy adults. J Appl Behav Anal 2014 Apr 17;47(2):231-245. [doi: 10.1002/jaba.119] [Medline: 24740477]
55. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
56. Zhang J, Iii JB. Mobile app-based small-group physical activity intervention for young African American women: a pilot randomized controlled trial. Prev Sci 2019 Aug 20;20(6):863-872. [doi: 10.1007/s11121-019-01006-4] [Medline: 30788692]
57. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating Mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
58. Butryn ML, Arigo D, Raggio GA, Colasanti M, Forman EM. Enhancing physical activity promotion in midlife women with technology-based self-monitoring and social connectivity: a pilot study. J Health Psychol 2016 Aug 10;21(8):1548-1555. [doi: 10.1177/1359105314558895] [Medline: 25488937]
59. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
60. Cadmus-Bertram L, Tevaarwerk AJ, Sesto ME, Gangnon R, Van Remortel B, Date P. Building a physical activity intervention into clinical care for breast and colorectal cancer survivors in Wisconsin: a randomized controlled pilot trial. J Cancer Surviv 2019 Aug 1;13(4):593-602 [FREE Full text] [doi: 10.1007/s11764-019-00778-6] [Medline: 31264183]
61. Duncan M, Fenton S, Brown W, Collins C, Glozier N, Kolt G, et al. Efficacy of a multi-component m-health weight-loss intervention in overweight and obese adults: a randomised controlled trial. Int J Environ Res Public Health 2020 Aug 26;17(17):6200 [FREE Full text] [doi: 10.3390/ijerph17176200] [Medline: 32859100]
62. Ellingson LD, Lansing JE, DeShaw KJ, Peyer KL, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019 Feb 14;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
63. Kandula NR, Dave S, De Chavez PJ, Marquez DX, Bharucha H, Mammen S, et al. An exercise intervention for South Asian mothers with risk factors for diabetes. Transl J Am Coll Sports Med 2016 Jun 15;1(6):52-59 [FREE Full text] [Medline: 27617303]
64. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. Obesity (Silver Spring) 2016 Aug 01;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
65. Singh B, Spence RR, Sandler CX, Tanner J, Hayes SC. Feasibility and effect of a physical activity counselling session with or without provision of an activity tracker on maintenance of physical activity in women with breast cancer - a randomised controlled trial. J Sci Med Sport 2020 Mar;23(3):283-290. [doi: 10.1016/j.jsams.2019.09.019] [Medline: 31640924]
66. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. BMC Cancer 2019 Mar 11;19(1):218 [FREE Full text] [doi: 10.1186/s12885-019-5427-5] [Medline: 30866859]
67. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
68. Robinson SA, Bisson AN, Hughes ML, Ebert J, Lachman ME. Time for change: using implementation intentions to promote physical activity in a randomised pilot trial. Psychol Health 2019 Feb 30;34(2):232-254 [FREE Full text] [doi: 10.1080/08870446.2018.1539487] [Medline: 30596272]
69. Schumacher LM, Arigo D, Thomas C. Understanding physical activity lapses among women: responses to lapses and the potential buffering effect of social support. J Behav Med 2017 Oct 5;40(5):740-749. [doi: 10.1007/s10865-017-9846-y] [Medline: 28382571]
70. Adams MA, Hurley JC, Todd M, Bhuiyan N, Jarrett CL, Tucker WJ, et al. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health 2017 Mar 29;17(1):1-16 [FREE Full text] [doi: 10.1186/s12889-017-4197-8] [Medline: 28356097]
71. Arigo D. Promoting physical activity among women using wearable technology and online social connectivity: a feasibility study. Health Psychol Behav Med 2015 Dec 31;3(1):391-409. [doi: 10.1080/21642850.2015.1118350]
72. Arigo D, Schumacher LM, Pinkasavage E, Butryn ML. Addressing barriers to physical activity among women: a feasibility study using social networking-enabled technology. Digit Health 2015 May 05;1:2055207615583564 [FREE Full text] [doi: 10.1177/2055207615583564] [Medline: 29942539]
73. Finkelstein J, Bedra M, Li X, Wood J, Ouyang P. Mobile app to reduce inactivity in sedentary overweight women. Stud Health Technol Inform 2015;216:89-92. [Medline: 26262016]
74. Fukuoka Y, Vittinghoff E, Hooper J. A weight loss intervention using a commercial mobile application in Latino Americans-Adelgaza Trial. Transl Behav Med 2018 Sep 08;8(5):714-723 [FREE Full text] [doi: 10.1093/tbm/ibx039] [Medline: 29474702]

75. Gell NM, Grover KW, Savard L, Dittus K. Outcomes of a text message, Fitbit, and coaching intervention on physical activity maintenance among cancer survivors: a randomized control pilot trial. J Cancer Surviv 2020 Feb 27;14(1):80-88. [doi: 10.1007/s11764-019-00831-4] [Medline: 31776849]
76. Gremaud AL, Carr LJ, Simmering JE, Evans NJ, Cremer JF, Segre AM, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. J Am Heart Assoc 2018 Jul 03;7(13):e007735. [doi: 10.1161/jaha.117.007735]
77. Grossman J, Arigo D, Bachman J. Meaningful weight loss in obese postmenopausal women: a pilot study of high-intensity interval training and wearable technology. Menopause 2018 Apr;25(4):465-470. [doi: 10.1097/GME.0000000000001013] [Medline: 29088015]
78. Linke SE, Hovsepians R, Schnebly B, Godfrey K, Noble M, Strong DR, et al. The Go-VAR (Veterans Active Recovery): an adjunctive, exercise-based intervention for veterans recovering from substance use disorders. J Psychoactive Drugs 2019 Jan 17;51(1):68-77. [doi: 10.1080/02791072.2018.1560518] [Medline: 30653409]
79. Meints SM, Yang HY, Collins JE, Katz JN, Losina E. Race differences in physical activity uptake within a workplace wellness program: a comparison of black and white employees. Am J Health Promot 2019 Jul 26;33(6):886-893 [FREE Full text] [doi: 10.1177/0890117119833341] [Medline: 30808208]
80. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
81. Reed J, Estabrooks P, Pozehl B, Heelan K, Wichman C. Effectiveness of the 5A's model for changing physical activity behaviors in rural adults recruited from primary care clinics. J Phys Act Health 2019 Dec 01;16(12):1138-1146. [doi: 10.1123/jpah.2018-0477] [Medline: 31553946]
82. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
83. Willis EA, Szabo-Reed AN, Ptomey LT, Steger FL, Honas JJ, Al-Hihi EM, et al. Distance learning strategies for weight management utilizing online social networks versus group phone conference call. Obes Sci Pract 2017 Jun 05;3(2):134-142 [FREE Full text] [doi: 10.1002/osp4.96] [Medline: 28713582]
84. Ashe MC, Winters M, Hoppmann CA, Dawes MG, Gardiner PA, Giangregorio LM, et al. "Not just another walking program": Everyday Activity Supports You (EASY) model-a randomized pilot study for a parallel randomized controlled trial. Pilot Feasibility Stud 2015 Jan 12;1(1):4 [FREE Full text] [doi: 10.1186/2055-5784-1-4] [Medline: 27175291]
85. Christiansen MB, Thoma LM, Master H, Voinier D, Schmitt LA, Ziegler ML, et al. Feasibility and preliminary outcomes of a physical therapist-administered physical activity intervention after total knee replacement. Arthritis Care Res (Hoboken) 2020 May 08;72(5):661-668. [doi: 10.1002/acr.23882] [Medline: 30908867]
86. Kenfield SA, Van Blarigan EL, Ameli N, Lavaki E, Cedars B, Paciorek AT, et al. Feasibility, acceptability, and behavioral outcomes from a technology-enhanced behavioral change intervention (Prostate 8): a pilot randomized controlled trial in men with prostate cancer. Eur Urol 2019 Jun;75(6):950-958. [doi: 10.1016/j.eururo.2018.12.040] [Medline: 30638635]
87. Rossi A, Frechette L, Miller D, Miller E, Friel C, Van Arsdale A, et al. Acceptability and feasibility of a Fitbit physical activity monitor for endometrial cancer survivors. Gynecol Oncol 2018 Jun;149(3):470-475. [doi: 10.1016/j.ygyno.2018.04.560] [Medline: 29692337]
88. Schmidt LI, Gabrian M, Jansen CP, Wahl HW, Sieverding M. Extending research on self-regulation of physical activity in older age: role of views on aging within an intensive ambulatory assessment scheme. J Self-regul Regul 2018;4:43-59. [doi: 10.11588/josar.2018.0.49362]
89. Streber A, Abu-Omar K, Hentschke C, Rütten A. A multicenter controlled study for dementia prevention through physical, cognitive and social activities - GESTALT-kompakt. Clin Interv Aging 2017 Dec;12:2109-2121 [FREE Full text] [doi: 10.2147/CIA.S141163] [Medline: 29276380]
90. Harkins KA, Kullgren JT, Bellamy SL, Karlawish J, Glanz K. A trial of financial and social incentives to increase older adults' walking. Am J Prev Med 2017 May;52(5):123-130. [doi: 10.1016/j.amepre.2016.11.011] [Medline: 28062271]
91. McMahon SK, Lewis B, Oakes JM, Wyman JF, Guan W, Rothman AJ. Assessing the effects of interpersonal and intrapersonal behavior change strategies on physical activity in older adults: a factorial experiment. Ann Behav Med 2017 Jun 10;51(3):376-390 [FREE Full text] [doi: 10.1007/s12160-016-9863-z] [Medline: 28188585]
92. Vidoni ED, Watts AS, Burns JM, Greer CS, Graves RS, Van Sciver A, et al. Feasibility of a memory clinic-based physical activity prescription program. J Alzheimer's Dis 2016 Jun 22;53(1):161-170. [doi: 10.3233/jad-160158]
93. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. JMIR Mhealth Uhealth 2015 Nov 19;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
94. Frechette L, Miller D, Rossi A, Miller E, Van Arsdale A, Viswanathan S, et al. Acceptability and feasibility of wearable fitness technology for endometrial cancer survivors. Gynecol Oncol 2018 Jun;149:223. [doi: 10.1016/j.ygyno.2018.04.505]

95. 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]
96. Chia GL, Anderson A, McLean LA. Behavior change techniques incorporated in fitness trackers: content analysis. JMIR Mhealth Uhealth 2019 Jul 23;7(7):e12768 [FREE Full text] [doi: 10.2196/12768] [Medline: 31339101]
97. Pumpera MA, Mendozaa JA, Arseniev-Koehlera A, Holma M, Waitea A, Morenoa MA. Using a Facebook group as an adjunct to a pilot ealth physical activity intervention: a mixed methods approach. In: Annual Review of Cybertherapy and Telemedicine. San Diego, California: Interactive Media Institute; 2015:97-101.
98. Goodyear VA, Kerner C, Quennerstedt M. Young people's uses of wearable healthy lifestyle technologies; surveillance, self-surveillance and resistance. Sport Educ Soc 2017 Sep 22;24(3):212-225. [doi: 10.1080/13573322.2017.1375907]
99. Kononova A, Li L, Kamp K, Bowen M, Rikard R, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
100. Ledger D, McCaffrey D. Inside wearables: how the science of human behavior change offers the secret to long-term engagement. Endeavour Partners. 2014. URL: https://archives.yegii.com/asset/inside-wearableshow-science-human-behavior-change-offers-secret-long-term-engagement-1513 [accessed 2021-05-07]
101. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
102. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/s2213-8587(16)30284-4]
103. Polgreen LA, Anthony C, Carr L, Simmering JE, Evans NJ, Foster ED, et al. The effect of automated text messaging and goal setting on pedometer adherence and physical activity in patients with diabetes: a randomized controlled trial. PLoS One 2018 May 2;13(5):e0195797 [FREE Full text] [doi: 10.1371/journal.pone.0195797] [Medline: 29718931]
104. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2018 Jul 08;52(13):844-850. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]
105. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]
106. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
107. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary intake and physical activity assessment: current tools, techniques, and technologies for use in adult populations. Am J Prev Med 2018 Oct;55(4):93-104 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.011] [Medline: 30241622]
108. Balbim G, Marques I, Marquez D, Patel D, Sharp L, Kitsiou S, et al. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. JMIR Mhealth Uhealth 2021 Mar 01;9(3):e25289 [FREE Full text] [doi: 10.2196/25289] [Medline: 33646135]
109. Ainsworth BE, Keller C, Herrmann S, Belyea M, Records K, Nagle-Williams A, et al. Physical activity and sedentary behaviors in postpartum Latinas: Madres para la Salud. Med Sci Sports Exerc 2013 Jul;45(7):1298-1306 [FREE Full text] [doi: 10.1249/MSS.0b013e3182863de5] [Medline: 23439416]
110. Owen N, Healy G, Matthews C, Dunstan D. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
111. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too little exercise and too much sitting: inactivity physiology and the need for new recommendations on sedentary behavior. Curr Cardiovasc Risk Rep 2008 Jul 17;2(4):292-298 [FREE Full text] [doi: 10.1007/s12170-008-0054-8] [Medline: 22905272]
112. Silfee VJ, Haughton CF, Jake-Schoffman DE, Lopez-Cepero A, May CN, Sreedhara M, et al. Objective measurement of physical activity outcomes in lifestyle interventions among adults: a systematic review. Prev Med Rep 2018 Sep;11:74-80 [FREE Full text] [doi: 10.1016/j.pmedr.2018.05.003] [Medline: 29984142]
113. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on trial registration 11 years after the ICMJE Policy was established. N Engl J Med 2017 Jan 26;376(4):383-391. [doi: 10.1056/nejmsr1601330]

Abbreviations

API: application programming interface
MVPA: moderate-to-vigorous physical activity
PA: physical activity

RCT: randomized controlled trial

Edited by L Buis; submitted 11.08.20; peer-reviewed by K Glanz, N Ridgers, J Alvarez Pitti; comments to author 22.10.20; revised version received 31.01.21; accepted 06.04.21; published 28.05.21.

Please cite as:
St Fleur RG, St George SM, Leite R, Kobayashi M, Agosto Y, Jake-Schoffman DE
Use of Fitbit Devices in Physical Activity Intervention Studies Across the Life Course: Narrative Review
JMIR Mhealth Uhealth 2021;9(5):e23411
URL: https://mhealth.jmir.org/2021/5/e23411
doi:10.2196/23411
PMID:34047705

Original Paper

On-site Dining in Tokyo During the COVID-19 Pandemic: Time Series Analysis Using Mobile Phone Location Data

Miharu Nakanishi[1,2], PhD; Ryosuke Shibasaki[3], PhD; Syudo Yamasaki[1], PhD; Satoshi Miyazawa[4], MPS; Satoshi Usami[5], PhD; Hiroshi Nishiura[6], PhD; Atsushi Nishida[1,7], PhD

[1]Research Center for Social Science & Medicine, Tokyo Metopolitan Institute of Medical Science, Setagaya-ku, Tokyo, Japan

[2]Department of Psychiatric Nursing, Tohoku University Graduate School of Medicine, Sendai-shi, Miyagi, Japan

[3]Division of Environmental Studies, Department of Socio-Cultural Environmental Studies, The University of Tokyo, Kashiwa-shi,Chiba, Japan

[4]Technology Department, LocationMind Inc, Chiyoda-ku, Tokyo, Japan

[5]Center for Research and Development of Higher Education, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

[6]School of Public Health, Kyoto University, Kyoto-shi, Kyoto, Japan

[7]Tokyo Center for Infectious Disease Control and Prevention, Shinjuku-ku, Tokyo, Japan

Corresponding Author:
Miharu Nakanishi, PhD
Research Center for Social Science & Medicine
Tokyo Metopolitan Institute of Medical Science
2-1-6 Kamikitazawa
Setagaya-ku, Tokyo, 156-8506
Japan
Phone: 81 3 6834 2292
Fax: 81 3 6834 2292
Email: mnakanishi-tky@umin.ac.jp

Abstract

Background: During the second wave of COVID-19 in August 2020, the Tokyo Metropolitan Government implemented public health and social measures to reduce on-site dining. Assessing the associations between human behavior, infection, and social measures is essential to understand achievable reductions in cases and identify the factors driving changes in social dynamics.

Objective: The aim of this study was to investigate the association between nighttime population volumes, the COVID-19 epidemic, and the implementation of public health and social measures in Tokyo.

Methods: We used mobile phone location data to estimate populations between 10 PM and midnight in seven Tokyo metropolitan areas. Mobile phone trajectories were used to distinguish and extract on-site dining from stay-at-work and stay-at-home behaviors. Numbers of new cases and symptom onsets were obtained. Weekly mobility and infection data from March 1 to November 14, 2020, were analyzed using a vector autoregression model.

Results: An increase in the number of symptom onsets was observed 1 week after the nighttime population volume increased (coefficient=0.60, 95% CI 0.28 to 0.92). The effective reproduction number significantly increased 3 weeks after the nighttime population volume increased (coefficient=1.30, 95% CI 0.72 to 1.89). The nighttime population volume increased significantly following reports of decreasing numbers of confirmed cases (coefficient=–0.44, 95% CI –0.73 to –0.15). Implementation of social measures to restaurants and bars was not significantly associated with nighttime population volume (coefficient=0.004, 95% CI –0.07 to 0.08).

Conclusions: The nighttime population started to increase after decreasing incidence of COVID-19 was announced. Considering time lags between infection and behavior changes, social measures should be planned in advance of the surge of an epidemic, sufficiently informed by mobility data.

(JMIR Mhealth Uhealth 2021;9(5):e27342) doi:10.2196/27342

KEYWORDS

COVID-19; mobility data; on-site dining; public health and social measures; public health; mobile phone; mobility; protection; time series; location; infectious disease; transmission

Introduction

As specific treatment and prevention measures have yet to be established, the COVID-19 global pandemic, caused by SARS-CoV-2, has required nations to implement public health and social measures, strategically focusing on mobility restrictions. SARS-CoV-2 is transmitted primarily by respiratory droplets via close face-to-face contact [1]. As the infection can be spread by asymptomatic and presymptomatic carriers, public health and social measures are warranted that target physical distancing and minimizing verbal interactions [2]. Many countries implemented national lockdowns in early 2020 that have had significant effects on reducing transmission [3]. Early and intensive interventions with lockdown periods are effective in reducing clinical cases and preventing an excess of health care demand compared with the supply [4]. However, once the lockdown measures were relaxed, second waves of COVID-19 gradually started, and large parts of Europe were forced to decide whether to implement second lockdowns [5]. National lockdowns and human mobility restrictions can potentially have immense adverse economic effects [6]. Governments worldwide now face the common challenge of easing lockdowns and restrictions while balancing out damages in various sectors, including health, social, and economic aspects [7]. Less restrictive public health and social measures to suppress the COVID-19 epidemic are called for, and such early interventions that are focused on reducing specific high risk behaviors are being explored.

Dining at restaurants, bars, and nightclubs involves removal of face masks and talking with others; thus, this activity is associated with elevated risk of COVID-19 transmission [8]. At a national level, in September 2020, restaurants, bars, and pubs in England were required to close at 10 PM [9]. Local measures to close bars and restaurants were also implemented in France from September 2020 [10]. In Japan, the national government's response to the second wave during July-August 2020 appeared to be slow, as the Go To Travel campaign that offered discounts on hotel charges and local coupons to encourage consumption was launched on July 22 [11]. The Tokyo Metropolitan Government asked restaurants and bars to close at 10 PM from August 3 to September 15, 2020 [12]. Tokyo metropolitan areas have not only the highest cumulative incidence of COVID-19 but also the highest population density [13], which is now recognized as the factor that determines the secondary transmission of COVID-19, and this area faced a second wave of COVID-19 starting in July 2020 [12]. To date, it has remained unclear how successfully public health and social measures can be implemented to reduce specific behaviors with high risk of infection (ie, on-site dining). Assessing the associations between human behavior, infection, and public health and social measures is therefore essential to understand achievable reductions and identify the factors driving the changes in social dynamics. The findings of such an assessment can have global implications for public health and social measures to suppress COVID-19.

Mobile phone location data can be used to monitor changes in human behavior across different locations in a country. These data indicate whether people are staying in the same location or moving around. The effects of public health and social measures to suppress COVID-19 transmission range from reduction of travel to distant locations to increased fractions of people staying at home, as assessed by mobile phone location data [14,15]. However, there is limited evidence regarding less restrictive COVID-19 countermeasures and social dynamics related to on-site dining. In this study, we used mobile phone trajectories to estimate the nighttime population volumes of people who stayed near restaurants and bars as a measure of on-site dining behavior. We investigated the association between nighttime population volumes, the COVID-19 epidemic, and the implementation of public health and social measures in Tokyo.

Methods

Timeline of the COVID-19 Pandemic in Tokyo

Infection data were collected during a 34-week period from March 1 to November 14, 2020. The study period was determined to observe the initial confirmed case reports in Tokyo, followed by the implementation of self-restraint–based contact reduction across all areas of Tokyo and social measures in restaurants and bars by the Tokyo Metropolitan Government. The number of new COVID-19 case reports and symptom onsets per day was made publicly available at websites by the Tokyo Metropolitan Government [16]. The number of symptom onsets was sequentially added over 3-week periods of prompt reports, as there were several time lags between symptom onset, testing, diagnosis, and inclusion in the daily report of new cases. We thus obtained data on the number of symptom onsets as of December 5, 2020. In this study, the weekly numbers of symptom onsets and confirmed case reports were used for analysis.

The effective reproduction number (R_t), the average number of secondary cases generated by a single primary case, was estimated as a function of the estimated calendar date of infection. To achieve this, the date of infection was statistically inferred using confirmed cases that were divided into two groups: cases with (1) known date of illness onset and (2) known date of confirmation and without known date of illness onset. For the latter group, we back-projected the suggested date of illness onset nonparametrically, and then we summed these cases with group (1). Subsequently, we implemented nonparametric back-projection of the date of infection using the incubation period distribution with a mean of 5.2 days. Using the assumed median generation time of 4.6 days [17], the renewal equation was employed to estimate everyday R_t. The Bayesian Markov chain Monte Carlo method was implemented using rstan, version 2.19.3; the estimation code is available in the GitHub repository [18]. To control for local heterogeneities that are exhibited as daily fluctuations, we smoothed the time series by computing the 7-day rolling averages, and we adopted the number on the fourth day of each week for analysis.

Mobile Phone Location Data

Population volumes at 10 PM to midnight in seven Tokyo metropolitan areas were estimated using location data from smartphones created by LocationMind xPop. LocationMind

xPop uses aggregated people flow data originally collected by NTT Docomo, Inc, through their application service "Docomo Map Navi" using only cell phone location data collected with user consent to the auto GPS function of the service, and the data are then statistically processed by NTT Docomo in their entirety before being provided to LocationMind Inc [19]. The original location data are GPS data (latitude and longitude) sent at a frequency of every 5 minutes at the shortest interval, and these data do not include information that specifies individuals [20]. NTT Docomo is the largest mobile phone operator in Japan, and it accounts for approximately 44% of total mobile phone subscribers [21]. The population per hour was estimated using the location data and calibration with the national census.

The users in this study agreed to provide their location information. The data are anonymized so that individuals cannot be specified, and personal information such as gender, age, and occupation are unknown. The population in this study represents the number of mobile devices that appear in a 500 m^2 grid at a specific time. Mobile phone trajectories were used to distinguish and extract on-site dining behavior from stay-at-work and stay-at-home behaviors. A stay point was assumed at which the GPS data of a device are concentrated within the area with positional errors for approximately 15 minutes or longer. A user's place of work and home were estimated using the information on the stay points, including period of time and length of stay per user. Stay points that were spatially distinct from the estimated place of work and home were accounted for with stay episodes other than stay-at-work and stay-at-home. The following seven Tokyo metropolitan areas were selected to represent districts with restaurants and bars: Kabuki-cho (mesh code 533945361), Ginza-Corridor-gai (mesh code 533946002), Shibuya-Center-gai (mesh code 533935952), Ueno-Nakamachi-dori (mesh code 533946512), Shinjuku-Ni-chome (mesh code 533945264), Ikebukuro (mesh code 533945764), and Roppongi (mesh code 533935984). The selection was based on designated areas in Tokyo for monitoring of people flow data by the Cabinet Office [22].

Nighttime Population Volume

Because using a single data source at a specific time may not represent the population distribution in the whole period, the average population concentration was used to estimate the nighttime population volumes in seven Tokyo metropolitan areas between 10 PM and midnight. A scaling factor was calculated using the estimated place of home for each user in combination with the national census, which gave a value of the total population represented by the user. The number of people staying in a certain area per hour was calculated by summing the scaling factors of the users. The estimated number of people per hour was calculated on a weekly basis to meet a sufficient sample size, as areas that contain few persons should be removed to prevent the identification of individuals based on their location.

Public Health and Social Measures

During the first wave of COVID-19, self-restraint–based contact reduction was implemented for the entire population of Tokyo for a 7-week period from April 7 to May 24, 2020. First-wave countermeasures included declaration of a state of emergency, call for self-restraint in travel and going out, restrictions of the maximum capacity of events and gatherings, and restrictions of use of facilities where clusters were occurring and the "3Cs" (closed space, crowded space, and closed contact setting) were observed [23]. Restaurants and bars were asked to cooperate by shutting down their services at 8 PM during this period.

During the second wave, social measures for restaurants and bars in Tokyo were in place for a 6-week period from August 3 to September 15, 2020. Restaurants and bars serving alcohol were asked to shorten their operating hours to 10 PM for this 6-week period.

Neither the first nor the second wave interventions were legally binding; for example, owners of restaurants and bars were asked to cooperate with the shutdown of services, and instead, the owner received a subsidy from Tokyo. In the analysis, the implementation of public health and social measures was presented as a dichotomous variable (implemented=1, absent=0).

Statistical Analysis

The initial and peak weeks were identified by visual inspection of the observed time series in the night-time population, symptom onsets, confirmed case reports, and R_t.

A vector autoregression (VAR) model was employed to predict endogenous outcomes (ie, nighttime population volume, symptom onsets, confirmed case reports, R_t) while investigating how they are influenced by each other in terms of Granger causality [24]. VAR is a system of equations with outcomes that depend on other outcome variables. Suppose we have a vector of time series data Wt, Xt, Yt, and Zt (four endogenous variables); then, the VAR model with one exogenous variable and p lags can be expressed as below:

$$Wt = \nu 1 + \rho 1Ct + \alpha 11Wt\text{-}1 + \ldots + \alpha 1pWt\text{-}p + \beta 11Xt\text{-}1 + \ldots + \beta 1pXt\text{-}p + \gamma 11Yt\text{-}1 + \ldots + \gamma 1pYt\text{-}p + \delta 11Zt\text{-}1 + \ldots + \delta 1pZt\text{-}p + \upsilon$$

$$Xt = \nu 2 + \rho 2Ct + \alpha 21Wt\text{-}1 + \ldots + \alpha 2pWt\text{-}p + \beta 21Xt\text{-}1 + \ldots + \beta 2pXt\text{-}p + \gamma 21Yt\text{-}1 + \ldots + \gamma 2pYt\text{-}p + \delta 21Zt\text{-}1 + \ldots + \delta 2pZt\text{-}p + \upsilon 2$$

$$Yt = \nu 3 + \rho 3Ct + \alpha 31Wt\text{-}1 + \ldots + \alpha 3pWt\text{-}p + \beta 31Xt\text{-}1 + \ldots + \beta 3pXt\text{-}p + \gamma 31Yt\text{-}1 + \ldots + \gamma 3pYt\text{-}p + \delta 31Zt\text{-}1 + \ldots + \delta 3pZt\text{-}p + \upsilon 3$$

$$Zt = \nu 4 + \rho 4Ct + \alpha 41Wt\text{-}1 + \ldots + \alpha 4pWt\text{-}p + \beta 41Xt\text{-}1 + \ldots + \beta 4pXt\text{-}p + \gamma 41Yt\text{-}1 + \ldots + \gamma 4pYt\text{-}p + \delta 41Zt\text{-}1 + \ldots + \delta 4pZt\text{-}p + \upsilon 4$$

where ν is the intercept, ρ is the coefficient of exogenous variable C, α is the coefficient of the lagged terms of W, β is the coefficient of the lagged terms of X, γ is the coefficient of the lagged terms of Y, δ is the coefficient of the lagged terms of Z, and υ is assumed to be white noise. In the W equation, β, γ, and δ infer how W is influenced by other outcomes (X, Y, and Z) in terms of Grander causality.

The number of lags can be statistically evaluated by information criteria such as the Akaike information criterion (AIC). To measure the effect of restrictions on growth in the nighttime population and infections, the first difference in the natural log was used for a respective time series to stabilize the variances

with removing trends, as was applied in a published study [25]. Regarding R_t, the first difference was used for analysis. First-wave countermeasures and social measures implemented by restaurants and bars were included in the VAR model as exogenous covariates. Of interest to our research was the effect of the implementation of restrictions on restaurants and bars on the nighttime population volume and the lagged coefficient of the nighttime population volume in symptom onsets and R_t. Granger causality tests were performed for each equation to assess whether any pairs among nighttime population volume, symptom onsets, confirmed case reports, and R_t had a causal relationship.

A sensitivity analysis was conducted using the vector autoregression model in which the symptom onsets and confirmed case reports were excluded from the endogenous variables.

All statistical analyses were performed using Stata, version 16.1 (StataCorp LLC). The significance level in the two-tailed tests was set to .05.

Results

The sum of the nighttime population volumes of the seven districts appeared to exhibit a decreasing trend before the implementation of first-wave countermeasures. During the self-restraint period, this sum started to increase, and it peaked around late June 2020. During the period of restrictions on restaurants and bars, the nighttime population volume consistently increased over time (Figure 1). The estimated volume of the nighttime population per area showed similar time trends from one place to the other (Multimedia Appendix 1). The first peak of confirmed case reports was observed in late March, and the second peak was around late July 2020 (Figure 1). Overall, the peak of confirmed case reports followed symptom onsets with a 3-week lag. The total number of symptom onsets during the study period was 22,902, accounting for 66% of confirmed case reports (n=34,467). The pattern of R_t during the study period was similar to that the of the nighttime population volume, with a decreasing trend before the implementation of first-wave countermeasures and social measures in restaurants and bars (Figure 2).

Figure 1. Nighttime population volumes in Tokyo metropolitan areas, number of new symptom onsets, and confirmed case reports of COVID-19 per week. LocationMind xPop © Location Mind Inc.

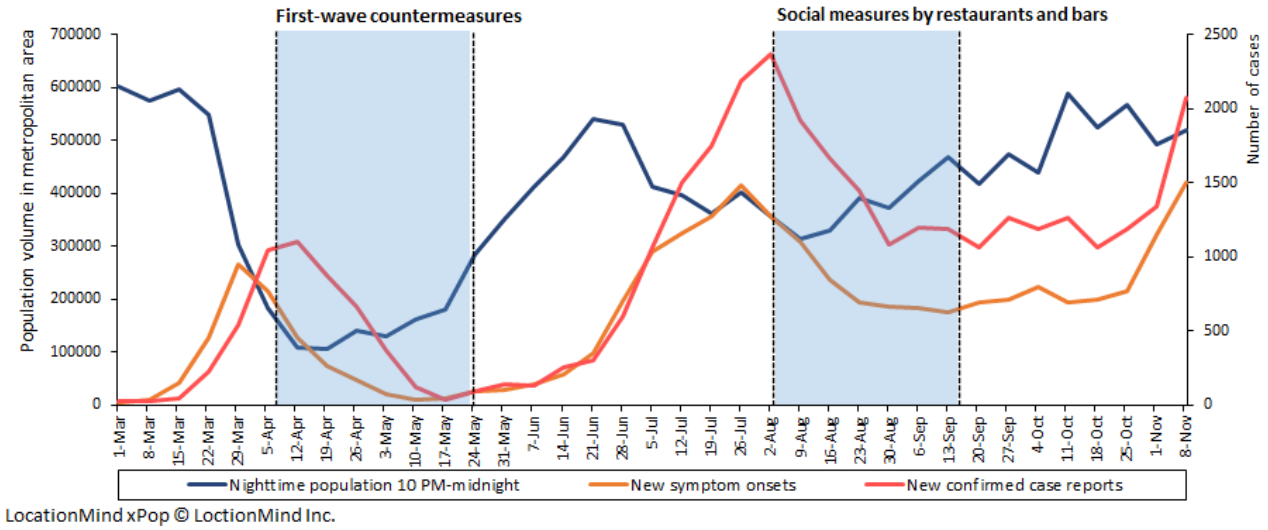
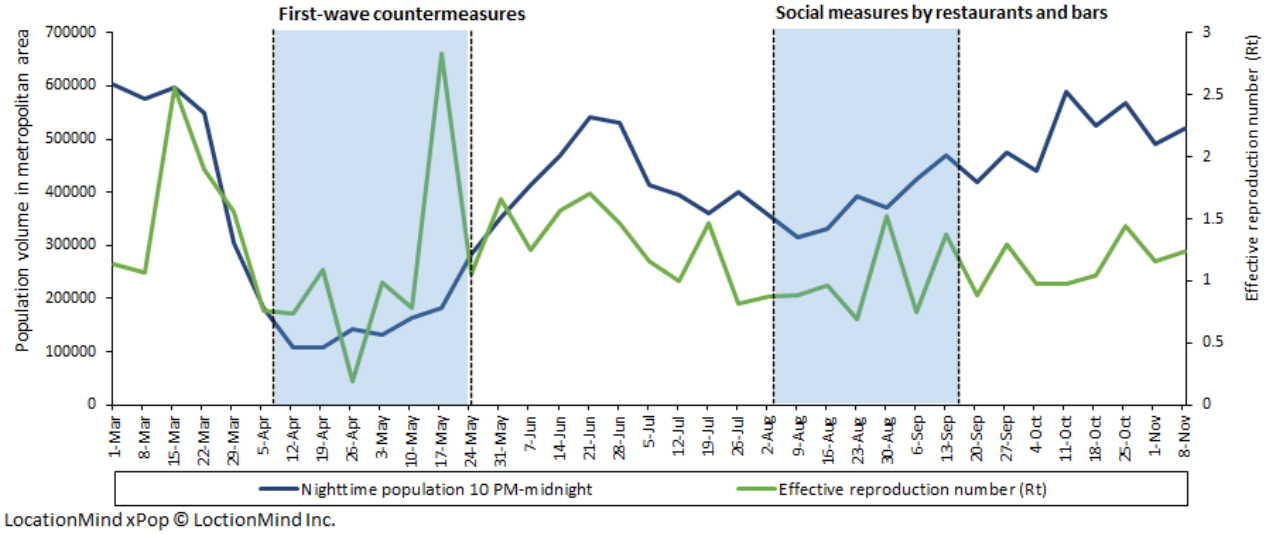


Figure 2. Nighttime population volumes in Tokyo metropolitan areas and the effective reproduction number of COVID-19 per week. The blue line presents the sum of the estimated numbers of people in the seven districts between 10 PM and midnight. The green line presents the effective reproduction number based on the number of confirmed cases in Tokyo. LocationMind xPop © Location Mind Inc.



For the lag-length selection of the VAR model, the Schwarz criterion, final prediction error, AIC, and Hannan-Quinn criterion suggest four lags. Consequently, we used four lags in the VAR model for further analysis. The results of the VAR model showed that the increase in R_t was significantly associated with increases in nighttime population volume 3 and 4 weeks prior (Table 1). The increase in symptom onset was significantly associated with increases in the nighttime population volume 1, 3, and 4 weeks prior. Additionally, the increase of the nighttime population volume was significantly associated with a decrease in confirmed case reports 1-4 weeks prior. The implementation of social measures in restaurants and bars was not significantly associated with the nighttime population volume. The implementation of lockdown was significantly associated with an increase in the nighttime population volume.

Table 1. Results of vector autoregression analysis on the number of confirmed case reports, symptom onsets, nighttime population volume, and R_t per week.[a]

Variable	Lag	R_t [b]		Nighttime population volume		Symptom onsets		Confirmed case reports	
		Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
R_t	1	–1.00 (–1.41 to –0.59)[c]	<.001	0.04 (–0.11 to 0.20)	.58	0.30 (0.06 to 0.55)[c]	.02	–0.03 (–0.22 to 0.17)	.80
	2	–0.17 (–0.76 to 0.42)	.58	–0.16 (–0.39 to 0.06)	.15	0.22 (–0.14 to 0.58)	.22	–0.24 (–0.52 to 0.05)	.10
	3	0.02 (–0.51 to 0.56)	.93	–0.71 (–0.91 to 0.50)*	<.001	0.12 (–0.20 to 0.44)	.48	–0.14 (–0.40 to 0.11)	.28
	4	–0.12 (–0.46 to 0.22)	.50	–0.38 (–0.51 to –0.25)[c]	<.001	0.07 (–0.13 to 0.28)	.48	0.08 (–0.09 to 0.24)	.36
Nighttime population volume	1	0.42 (–0.11 to 0.95)	.12	–0.25 (–0.45 to –0.05)[c]	.01	0.60 (0.28 to 0.92)[c]	<.001	–0.05 (–0.31 to 0.20)	.69
	2	0.41 (–0.14 to 0.95)	.14	0.29 (0.08 to 0.49)[c]	.006	–0.17 (–0.50 to 0.16)	.30	–0.26 (–0.52 to –0.001)[c]	.05
	3	1.30 (0.72 to 1.89)[c]	<.001	–0.52 (–0.74 to –0.29)[c]	<.001	0.63 (0.28 to 0.99)[c]	<.001	–0.33 (–0.61 to –0.05)[c]	.02
	4	0.59 (0.04 to 1.14)[c]	.03	–0.03 (–0.24 to 0.18)	.77	0.59 (0.26 to 0.92)[c]	<.001	0.07 (–0.20 to 0.33)	.62
Symptom onset	1	–0.57 (–1.25 to 0.11)	.10	0.02 (–0.24 to 0.27)	.90	0.35 (–0.06 to 0.76)	.09	1.40 (1.07 to 1.72)[c]	<.001
	2	–1.50 (–2.69 to –0.32)[c]	.01	1.05 (0.61 to 1.50)[c]	<.001	–0.47 (–1.19 to 0.25)	.20	0.15 (–0.42 to 0.71)	.62
	3	1.42 (0.50 to 2.35)[c]	.002	–0.10 (–0.45 to 0.25)	.59	0.12 (–0.44 to 0.68)	.68	–0.35 (–0.79 to 0.09)	.12
	4	–0.58 (–1.33 to 0.18)	.13	0.38 (0.10 to 0.67)[c]	.008	0.24 (–0.22 to 0.70)	.31	0.45 (0.09 to 0.81)[c]	.02
Confirmed case reports	1	0.14 (–0.62 to 0.91)	.72	–0.44 (–0.73 to –0.15)[c]	.003	0.06 (–0.40 to 0.53)	.79	–0.37 (–0.74 to –0.003)[c]	.048
	2	–0.20 (–0.80 to 0.39)	.50	–0.29 (–0.52 to –0.07)[c]	.01	0.25 (–0.11 to 0.61)	.17	0.02 (–0.27 to 0.30)	.91
	3	–0.22 (–0.70 to 0.26)	.36	–0.65 (–0.84 to –0.47)[c]	<.001	–0.22 (–0.51 to 0.07)	.14	–0.07 (–0.30 to 0.16)	.55
	4	0.51 (0.18 to 0.84)[c]	.002	–0.31 (–0.44 to –0.19)[c]	<.001	0.07 (–0.12 to 0.27)	.46	–0.18 (–0.34 to –0.02)[c]	.02
Social measures in restaurants and bars (August 3 to September 15, 2020)	N/A[d]	–0.07 (–0.27 to 0.14)	.52	0.004 (–0.07 to 0.08)	.91	–0.07 (–0.20 to 0.05)	.25	–0.02 (–0.11 to 0.08)	.72
First-wave countermeasures (April 7 to May 24, 2020)	N/A	0.04 (–0.29 to 0.38)	.82	0.18 (0.05 to 0.30)[c]	.006	–0.03 (–0.24 to 0.17)	.74	–0.05 (–0.21 to 0.11)	.56

[a]The number of optimal lags was determined by information criteria, including the Akaike information criterion, Hannan–Quinn information criterion, Schiwarz-Bayesian information criteria, and final prediction error.

[b]R_t: effective reproduction number.

[c]Significant at $P<.05$.

[d]N/A: not applicable.

The results of the Granger causality tests indicated the existence of bidirectional causality running from confirmed case reports to the nighttime population, bidirectional causality running from the nighttime population to symptom onsets, and bidirectional causality running from the nighttime population to R_t (Table 2).

Table 2. Granger causality tests of pairs between nighttime population, symptom onsets, confirmed case reports, and R_t.

Equation	Excluded	Chi-square (*4*)	*P* value
R_t [a]	Nighttime population	42.36	<.001
	Symptom onsets	18.12	.001
	Confirmed case reports	13.67	.008
Nighttime population	R_t	67.28	<.001
	Symptom onsets	26.45	<.001
	Confirmed case reports	62.51	<.001
Symptom onsets	R_t	9.42	.051
	Nighttime population	47.71	<.001
	Confirmed case reports	10.36	.04
Confirmed case reports	R_t	14.39	.006
	Nighttime population	12.93	.01
	Symptom onsets	82.23	<.001

[a]R_t: effective reproduction number.

The lagged coefficients of the nighttime population to R_t remained significant (1 week prior, coefficient 1.08, 95% CI 0.28-1.87; 2 weeks prior, coefficient 0.90, 95% CI 0.10-1.70) in the sensitivity analysis, in which symptom onsets and confirmed case reports were excluded from the endogenous variables. The bidirectional causality running from the night-time population to R_t was also significant (χ^2_4= 27.00, *P*<.001) in this analysis.

Discussion

Principal Results

The increase in the nighttime population in the seven examined districts in Tokyo was followed by increased symptom onsets and increased R_t thereafter. Even during the intervention period, the nighttime population increased significantly following the decrease in confirmed case reports. The implementation of social measures by restaurants and bars was not significantly associated with the nighttime population during the second wave in Tokyo.

The results of our study imply that people adjusted their level of mobility according to the number of confirmed case reports and, thus, their perceived risk of their own infection. They may have reduced on-site dining behavior voluntarily after an increasing number of new confirmed cases was reported and then relaxed their mobility restrictions after a decreasing number of cases was reported. Although people dynamically adjusted their behavior in response to information and policies [26] and recognized that dining inside restaurants should be avoided [27], public attitudes related to behavioral restrictions changed as the COVID-19 epidemic changed. It is notable that there were considerable time lags between behavior changes, symptom onsets, and R_t. A time lag between major mobility reduction and the peak of confirmed cases was also observed in South Korea [28]. There could have been a delay in the voluntary reduction in on-site dining from the time that suppression of the surge of COVID-19 should have begun. This latency was also observed in the reaction to policy measures to restrict mobility in northern Italy [29]. Therefore, policy measures to mitigate COVID-19 should be planned at an earlier stage informed by social behavior dynamics, rather than based on the number of confirmed case reports. Early implementation of strategies is crucial to successful COVID-19 suppression and to avoid national lockdowns [30]. A tracking system of the nighttime population using mobile phone location data would be helpful for policy decision makers to monitor these dynamics in a real-time manner. Space-time dispersions in transmission of COVID-19 have occurred from metropolitan areas towards the countryside [31]. Thus, an automated information system to support strategic policy decision making is of particular importance in metropolitan areas with high population density and mobility, where there is an elevated risk of COVID-19 transmission.

The implementation of first-wave interventions in early 2020 was significantly associated with an increase in the nighttime population volume in Tokyo. Less restrictive measures, requesting that restaurants and bars shorten their operating hours, were not significantly associated with the nighttime population volume. The nighttime population volume started to decrease one month before the implementation of first-wave countermeasures and social measures to restaurants and bars, and it reached the minimum in the early phase of the interventions. People may have reduced on-site dining behavior earlier than the implementation of public health and social measures in response to preceding policy arguments and announcement of the implementations along with the increase

in confirmed cases. It should be noted that the decreasing trend appeared to be weaker against the second wave than in March 2020. Although we could not confirm this from our data, many people may have experienced caution fatigue by repeated COVID-19 surges and the highest number of new cases, and they may have become tired of physical distancing. Another possible explanation is that the public health and social measures were suboptimal to suppress the recovery of on-site dining behavior based on the level of the restrictions and the timing of the implementation. However, we could not determine the optimal level of restrictions and length of interventions with regard to COVID-19 mitigation or the economic effects of the interventions. Closure of restaurants and bars may result in decreased consumption expenditure and economic decline. Further investigation is needed to estimate the effect of reductions in the nighttime population volume on the economy and on COVID-19 transmission. This estimation can also inform the government as to how financial compensation should be provided to entities that comply with requests to close at night. In Japan, the national government launched a Go To Eat campaign to encourage people to dine at restaurants starting in October 2020 [32]. This campaign aimed to support the recovery of the industry from economic decline caused by first-wave countermeasures and other behavior restrictions; however, there are concerns that it increased on-site dining and that it resulted in the recent large resurgence of COVID-19 [33]. Strategies to support businesses need to be organized in conjunction with a level of behavior restrictions to prevent COVID-19 resurgence.

Limitations

This is the first study to relate public health and social measures to prevent COVID-19 transmission to the nighttime population volume in Tokyo. A strength of this study is the estimation of probable on-site dining behavior as opposed to stay-at-work and stay-at-home behavior using mobile phone location data. However, our study has some limitations that must be noted. The suspected mode of transmission in each case was unavailable due to the nature of open data, potentially leading to underestimation of the association between the nighttime population volume and the onset of symptoms. Lags may have been shorter if symptom onset were stratified by transmission via on-site dining. The risk of COVID-19 transmission in on-site dining could also be mitigated over time as restaurants and bars increasingly adopted physical distancing guidelines [34]. The number of reported cases in the first wave might have been suppressed because of limited testing capacity in early 2020. The effectiveness of the measures to reduce on-site dining may vary across regions and countries that have different response levels to policies among the general population. The level of effectiveness would also differ according to the stage of pandemic (first, second, or later), as public health and social measures against later waves can be subject to cumulative caution fatigue against COVID-19. Social behavioral dynamics could also be caused by the media's reporting on COVID-19 [35]. Future examination on the association between on-site dining and media reporting will be beneficial to provide guidelines to the media that support effective mobility restrictions.

Conclusions

Regardless of restrictions on restaurants and bars in Tokyo metropolitan areas, people reduced their mobility restrictions after decreases in new confirmed cases were reported. An increase in the nighttime population volume may result in increased symptom onset and R_t. A tracking system of the nighttime population using mobile phone location data would be helpful for policy decision makers to monitor social behavior dynamics and suppress COVID-19 transmission at an earlier stage.

Acknowledgments

The authors thank Dr Mitsuo Kaku (Tohoku Medical and Pharmaceutical University) and Tokyo Center for Infectious Disease Control and Prevention for their advice. The authors received financial support from the Tokyo Metropolitan Institute of Medical Science, Japan. HN received financial support from the World Health Organization and Japan Agency for Medical Research and Development (JP20fk0108140) for this research. The funders of this study had no role in the study design, data analysis, data interpretation, or writing of the article.

Authors' Contributions

MN, RS, and AN conceived and designed the study. MN, SM, and SU analyzed the data and performed the analysis. RS, HN, SU, SY, and AN interpreted the results. MN drafted the article. All authors contributed to writing the final version of the article.

Conflicts of Interest

SM is an employee and a shareholder of LocationMind Inc.

Multimedia Appendix 1
Estimated numbers of the nighttime population in seven Tokyo metropolitan areas between 10 PM and midnight per week from April 7 to May 24, 2020, in Tokyo.
[PNG File , 53 KB - mhealth_v9i5e27342_app1.png]

References

1. Wiersinga W, Rhodes A, Cheng A, Peacock S, Prescott H. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 2020 Aug 25;324(8):782-793 [FREE Full text] [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
2. Ng O, Marimuthu K, Koh V, Pang J, Linn K, Sun J, et al. SARS-CoV-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study. Lancet Infect Dis 2021 Mar;21(3):333-343. [doi: 10.1016/s1473-3099(20)30833-1]
3. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug;584(7820):257-261 [FREE Full text] [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
4. Davies N, Kucharski A, Eggo R, Gimma A, Edmunds W, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 2020 Jul;5(7):e375-e385 [FREE Full text] [doi: 10.1016/S2468-2667(20)30133-X]
5. Gurdasani D, Bear L, Bogaert D, Burgess R, Busse R, Cacciola R, et al. The UK needs a sustainable strategy for COVID-19. Lancet 2020 Dec;396(10265):1800-1801 [FREE Full text] [doi: 10.1016/s0140-6736(20)32350-3]
6. Bonaccorsi G, Pierri F, Cinelli M, Flori A, Galeazzi A, Porcelli F, et al. Economic and social consequences of human mobility restrictions under COVID-19. Proc Natl Acad Sci U S A 2020 Jul 07;117(27):15530-15535 [FREE Full text] [doi: 10.1073/pnas.2007658117] [Medline: 32554604]
7. Han E, Tan M, Turk E, Sridhar D, Leung G, Shibuya K, et al. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. Lancet 2020 Nov;396(10261):1525-1534 [FREE Full text] [doi: 10.1016/s0140-6736(20)32007-9]
8. Fisher K, Tenforde M, Feldstein L, Lindsell CJ, Shapiro NI, Files DC, CDC COVID-19 Response Team. Community and close contact exposures associated with COVID-19 among symptomatic adults ≥18 years in 11 outpatient health care facilities - United States, July 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1258-1264 [FREE Full text] [doi: 10.15585/mmwr.mm6936a5] [Medline: 32915165]
9. Coronavirus (COVID-19): what has changed? 22 September. UK Cabinet Office. 2020 Sep 22. URL: https://www.gov.uk/government/news/coronavirus-covid-19-what-has-changed-22-september [accessed 2020-11-20]
10. Coronavirus: Paris to shut bars and raise alert to maximum. BBC. 2020 Oct 05. URL: https://www.bbc.com/news/world-europe-54413563 [accessed 2020-11-20]
11. Anzai A, Nishiura H. "Go To Travel" campaign and travel-associated coronavirus disease 2019 cases: a descriptive analysis, July-August 2020. J Clin Med 2021 Jan 21;10(3):398 [FREE Full text] [doi: 10.3390/jcm10030398] [Medline: 33494315]
12. Disaster prevention information. Webpage in Japanese. Tokyo Metropolitan Government. URL: https://www.bousai.metro.tokyo.lg.jp/1009757/1010005.html [accessed 2020-11-20]
13. Population and Households of Japan (Final Report of the 2015 Population Census). Statistics Bureau of Japan. URL: http://www.stat.go.jp/english/data/kokusei/2015/final_en/final_en.html [accessed 2021-02-07]
14. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. Lancet Digital Health 2020 Dec;2(12):e638-e649. [doi: 10.1016/s2589-7500(20)30243-0]
15. Arimura M, Ha T, Okumura K, Asada T. Changes in urban mobility in Sapporo city, Japan due to the Covid-19 emergency declarations. Trans Res Interdiscip Perspect 2020 Sep;7:100212. [doi: 10.1016/j.trip.2020.100212]
16. Updates on COVID-19 in Tokyo. Webpage in Japanese. Tokyo Metropolitan Government. URL: https://stopcovid19.metro.tokyo.lg.jp/ [accessed 2020-11-21]
17. Nishiura H, Linton N, Akhmetzhanov A. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020 Apr;93:284-286 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.060] [Medline: 32145466]
18. Jung S, Akhmetzhanov A, Mizumoto K, Nishiura H. Real-time estimation of the effective reproduction number of COVID-19 in Japan. GitHub. URL: https://github.com/contactmodel/COVID19-Japan-Reff [accessed 2021-04-30]
19. Location Mind xPop. Location Mind Inc. URL: https://locationmind.com/products/xpop-e/ [accessed 2021-05-03]
20. Xia T, Song X, Zhang H, Song X, Kanasugi H, Shibasaki R. Measuring spatio-temporal accessibility to emergency medical services through big GPS data. Health Place 2019 Mar;56:53-62 [FREE Full text] [doi: 10.1016/j.healthplace.2019.01.012] [Medline: 30703630]
21. Number of subscribers by carriers. Webpage in Japanese. Telecommunications Carriers Association. URL: https://www.tca.or.jp/database/ [accessed 2021-02-07]
22. COVID-19 information and resources. Webpage in Japanese. Cabinet Secretariat. URL: https://corona.go.jp/dashboard/ [accessed 2021-02-07]
23. Basic policies for novel coronavirus disease control by the Government of Japan (summary). Government of Japan. 2020 Mar 28. URL: https://www.mhlw.go.jp/content/10900000/000634753.pdf [accessed 2020-11-30]
24. Granger C. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 1969 Aug;37(3):424-438. [doi: 10.2307/1912791]

25. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020 Aug;584(7820):262-267 [FREE Full text] [doi: 10.1038/s41586-020-2404-8] [Medline: 32512578]
26. Cho S. Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: the case of Sweden. Econ J 2020;23(3):323-344. [doi: 10.1093/ectj/utaa025]
27. Czeisler M, Tynan MA, Howard ME, Honeycutt S, Fulmer EB, Kidder DP, et al. Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance - United States, New York City, and Los Angeles, May 5-12, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):751-758. [doi: 10.15585/mmwr.mm6924e1] [Medline: 32555138]
28. Yoon Y, Oh S, Cho J, Shin Y, Kim S, Kim N, et al. The effect of adaptive mobility policy to the spread of COVID-19 in urban environment: intervention analysis of Seoul, South Korea. ArXiv. Preprint posted online on September 16, 2020. [FREE Full text]
29. Piccinini D, Giunchi C, Olivieri M, Frattini F, Di Giovanni M, Prodi G, et al. COVID-19 lockdown and its latency in Northern Italy: seismic evidence and socio-economic interpretation. Sci Rep 2020 Oct 05;10(1):16487. [doi: 10.1038/s41598-020-73102-3] [Medline: 33020508]
30. Summers S, Cheng H, Lin H, Barnard D, Kvalsvig D, Wilson P, et al. Potential lessons from the Taiwan and New Zealand health responses to the COVID-19 pandemic. Lancet Regional Health - Western Pacific 2020 Nov;4:100044. [doi: 10.1016/j.lanwpc.2020.100044]
31. Gomes D, Andrade L, Ribeiro C, Peixoto M, Lima S, Duque A, et al. Risk clusters of COVID-19 transmission in northeastern Brazil: prospective space–time modelling. Epidemiol Infect 2020 Aug 24;148:e188 [FREE Full text] [doi: 10.1017/s0950268820001843]
32. Shimizu K, Kondo T, Tokuda Y, Shibuya K. An open letter to Japan's new Prime Minister. Lancet 2020 Oct;396(10259):e57. [doi: 10.1016/s0140-6736(20)32033-x]
33. Shimizu K, Wharton G, Sakamoto H, Mossialos E. Resurgence of covid-19 in Japan. BMJ 2020 Aug 18;370:m3221 [FREE Full text] [doi: 10.1136/bmj.m3221] [Medline: 32816821]
34. Gursoy D, Chi C. Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda. J Hosp Mark Manag 2020 Jul 05;29(5):527-529 [FREE Full text] [doi: 10.1080/19368623.2020.1788231]
35. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7 [FREE Full text] [doi: 10.1017/s003329172000224x]

Abbreviations

AIC: Akaike information criterion
R_t: effective reproduction number
VAR: vector autoregression

Edited by G Eysenbach; submitted 22.01.21; peer-reviewed by K Shimizu, H Kadir, S Pesälä; comments to author 06.02.21; revised version received 10.02.21; accepted 11.04.21; published 11.05.21.

Please cite as:
Nakanishi M, Shibasaki R, Yamasaki S, Miyazawa S, Usami S, Nishiura H, Nishida A
On-site Dining in Tokyo During the COVID-19 Pandemic: Time Series Analysis Using Mobile Phone Location Data
JMIR Mhealth Uhealth 2021;9(5):e27342
URL: https://mhealth.jmir.org/2021/5/e27342
doi:10.2196/27342
PMID:33886486

Original Paper

Smartphone and Tablet Software Apps to Collect Data in Sport and Exercise Settings: Cross-sectional International Survey

Matthew Peter Shaw[1*], MSc; Liam Paul Satchell[2*], PhD; Steve Thompson[3], PhD; Ed Thomas Harper[4], MSc; Carlos Balsalobre-Fernández[5], PhD; Daniel James Peart[6], PhD

[1]Sports, Physical Activity and Food, Western Norway University of Applied Sciences, Sogndal, Norway

[2]Department of Psychology, University of Winchester, Winchester, United Kingdom

[3]Academy of Sport and Physical Activity, Sheffield Hallam University, Sheffield, United Kingdom

[4]Number One Strength and Performance, Deeside, United Kingdom

[5]Applied Biomechanics and Sports Technology Research Group, Universidad Autónoma de Madrid, Madrid, Spain

[6]Department of Sport, Exercise and Rehabilitation, Northumbria University, Newcastle, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Matthew Peter Shaw, MSc
Sports, Physical Activity and Food
Western Norway University of Applied Sciences
Røyrgata 6
Sogndal, 6856
Norway
Phone: 47 57676391
Email: matthew.shaw@hvl.no

Abstract

Background: Advances in smartphone technology have facilitated an increase in the number of commercially available smartphone and tablet apps that enable the collection of physiological and biomechanical variables typically monitored in sport and exercise settings. Currently, it is not fully understood whether individuals collect data using mobile devices and tablets, independent of additional hardware, in their practice.

Objective: This study aims to explore the use of smartphone and tablet software apps to collect data by individuals working in various sport and exercise settings, such as sports coaching, strength and conditioning, and personal training.

Methods: A total of 335 practitioners completed an electronic questionnaire that surveyed their current training practices, with a focus on 2 areas: type of data collection and perceptions of reliability and validity regarding app use. An 18-item questionnaire, using a 5-point Likert scale, evaluated the perception of app use.

Results: A total of 204 respondents reported using apps to directly collect data, with most of them (196/335, 58.5%) collecting biomechanical data, and 41.2% (138/335) respondents reported using at least one evidence-based app. A binomial general linear model determined that evidence accessibility (β=.35, 95% CI 0.04-0.67; *P*=.03) was significantly related to evidence-based app use. Age (β=−.03, 95% CI −0.06 to 0.00; *P*=.03) had a significant negative effect on evidence-based app use.

Conclusions: This study demonstrates that practitioners show a greater preference for using smartphones and tablet devices to collect biomechanical data such as sprint velocity and jump performance variables. When it is easier to access information on the quality of apps, practitioners are more likely to use evidence-based apps. App developers should seek independent research to validate their apps. In addition, app developers should seek to provide clear signposting to the scientific support of their software in alternative ways.

(JMIR Mhealth Uhealth 2021;9(5):e21763) doi:10.2196/21763

KEYWORDS

mobile apps; sports; smartphone; mobile phone; questionnaire; survey

Introduction

Advances in smartphone technology have facilitated an increase in the number of commercially available smartphone and tablet apps, enabling the collection of various physiological and biomechanical variables without additional hardware. Smartphones and tablets typically contain a microphone, camera, light sensor, accelerometer, gyroscope, inclinometer, and magnetometer. These hardware components are used in combination with software apps to provide a variety of measurements. Several studies have demonstrated that smartphone and tablet cameras can validly and reliably measure biomechanical variables such as sprint time [1], movement velocity [2,3], and jump height [4]. Both the accelerometer [5] and magnetometer [6] have been used to validly and reliably measure the range of motion in multiple joints, and the inclinometer has been shown to validly and reliably determine break-point angle in the Nordic hamstring exercise [7]. In relation to light sensor hardware, Coppetti et al [8] have examined the ability of smartphones and tablets to measure heart rate via photoplethysmography.

Using smartphones and tablets, sport scientists and coaches are now able to collect data in practical settings, such as during match play or training, more economically. Compared with specialized hardware, commercially available smartphone apps are available at much lower costs or completely free of charge. For example, Romero-Franco et al [1] demonstrated that a smartphone app had comparable reliability and validity to timing gates, costing approximately 400 times more. Mobile technology has the potential to address problems with portability, cost, and time that are historically associated with laboratory-based equipment. However, using smartphones and tablets to measure physiological variables should be performed with caution. For example, there is some inconsistency in the use of software apps to measure heart rate [8], particularly during exercise of various intensities [9]. Furthermore, a recent review by Peart et al [10] highlighted inconsistencies in validity and reliability when estimating body fat percentage using a range of commercially available software [11-13].

Currently, there is very limited existing research that has investigated the use of smartphone or tablet software apps to collect data in sport and exercise settings. Most recently, Bromilow et al [14] surveyed exercise professionals in Australia to examine smartphone use in practice, concluding that smartphone use is highly prevalent in sport and exercise settings, but this is typically for tracking variables. *Tracking* is a term consistent in the sport and exercise literature [15], which refers to apps and software available to log training information. This can include, for example, running distance, resistance exercise repetitions, and heart rate. Typically, users enter this information themselves. Extending on the recent work of Bromilow et al [14], this investigation is the first study that primarily focuses on how practitioners use, or do not use, software apps that use mobile device hardware to collect data directly from the primary source. Furthermore, the existing literature demonstrates inconsistencies in validity and reliability, depending on the type of variable collected, and practitioners should, therefore, be critical in their selection of apps used to collect data. We currently do not fully understand whether practitioners collect data using mobile devices or if they do so using valid and reliable apps. Therefore, the primary aim of this investigation is to explore practitioners' use of smartphone or tablet software apps for collecting data. A secondary aim is to examine if practitioners select valid and reliable software apps and what may influence this selection.

Methods

Overview

An exploratory descriptive study was conducted to examine the current use of smartphones and tablets in sport and exercise settings, providing detailed information on practitioners' use of this technology. The study used a multiple-choice questionnaire survey that generated exploratory descriptive statistics of app use. The open survey was electronic, with links to the survey distributed using social media platforms such as Twitter and Facebook. The use of Facebook included distributing the survey in specialist groups such as the various National Strength and Conditioning Association special interest groups. Each coauthor sent emails with the links to the survey within their respective professional networks. A final strategy included advertising the survey during international conference presentations. The study procedure was approved by the institutional ethics committee of Sheffield Hallam University (ER8496574) in accordance with the seventh revision of the Declaration of Helsinki in advance of data collection. Survey responses were stored in password-protected files, where only the investigation team could access them.

Participants

Participants were required to be older than 18 years and engaged in the sport and exercise industry in either an employed or voluntary role to meet our inclusion criteria. Roles included sport scientist, strength and conditioning coach, physical education teacher, sports coach, and personal trainer. Before completing the questionnaire, participants were directed to the participant information section and informed of their right to withdraw from the study. Participation was voluntary, and no incentives were provided. Respondents were required to provide complete responses to the questions. If a question was not answered, they were not able to move to the subsequent question. Valid consent was obtained if the questionnaire was completed [16].

Survey Instrument

The survey consisted of a series of multiple-choice questions. The survey was developed using Google Forms, allowing participants to complete it remotely, and responses were automatically captured. The survey was open to any visitor to the survey URL. The questions differed based on the previous responses given. The survey gathered 3 areas of information:

- Demographic information: age, gender, and country of residence
- Industrial experience: area of study, area of employment, years of experience, vocational training, professional accreditation, and populations worked with

- App use: type of data collection (eg, cardiovascular), hardware used, and perceptions of reliability and validity.

Participants were given the opportunity to list all the apps they were currently using in practice. Once the survey was closed for responses, we reviewed all reported apps to identify those that had existing literature evidencing reliability and validity. To do so, a series of searches were conducted in Google Scholar and PubMed using the name of each reported app. Only the name of the app was included to determine if it was featured in any existing literature. Once an app was found in any experimental study, we assessed whether the authors reported an app to demonstrate evidence of both validity and reliability. In this investigation, the term *evidence-based app* refers to any app a respondent reported to use that had peer-reviewed evidence of acceptable reliability and validity.

Participants were required to complete an 18-item questionnaire (Textbox 1) to evaluate their perception of app use. The 18 items were formed following consultation with a panel of sport scientists who were independent of the authors' team and had expertise in survey design. A draft survey structure went through 2 rounds of feedback, following panel feedback and agreement between the coauthors. All items used a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree). These 18-items were generated for the target themes of reliability, validity, cost, and ease of use.

Textbox 1. Smartphone or tablet app perception questionnaire.

Q1. The apps I use in practice are difficult to use
Q2. Reliability of the apps is important
Q3. It is difficult to determine the validity of the apps I use
Q4. The price is important for me when selecting an app
Q5. I do not deem reliability of apps to be important
Q6. Other equipment is harder to use than apps for collecting the same data
Q7. It is easy to determine the reliability of the apps I use
Q8. I am more likely to use an app I have to pay for
Q9. It is easy to determine the validity of the apps I use
Q10. I do not consider the validity of apps to be important
Q11. The apps I use in practice are easy to use
Q12. It is important that there are reliability studies available for the apps I use
Q13. I am more likely to use an app if it is free
Q14. The price is not important for me when selecting an app
Q15. It is important that there are validity studies available for the apps I use
Q16. It is difficult to determine the reliability of the apps I use
Q17. It is important that apps have validity
Q18. Other equipment is easier to use than apps for collecting the same data

Statistical Analysis

As many of the survey items could be measuring the same core construct, the patterns of responses to the questionnaire were examined using exploratory factor analysis (EFA). Due to missing responses to the questionnaire, 261 cases were used for the EFA. The R statistical package *jmv* was used [17] to conduct an oblique (*oblimin*) minimum residual method EFA. Parallel analysis of the ascending number of factor models suggested that a 4-factor model fit was most effective, explaining a cumulative 36.91% of the variance, with adequate fit indices (Root mean square error of approximation=0.07; Tucker-Lewis index=0.86; χ^2_{87}=195.0; P<.001). Factor membership of items was assigned based on the strongest loading of an item onto a factor, with all loadings being at least stronger than 0.30.

The 4 factors that emerged in the data were evidence availability, evidence accessibility, nonevidence use, and resources. Table 1 shows the factor loadings with the strongest loading factor highlighted in italics. It is worth noting that the question, "I am more likely to use an app I have to pay for," poorly loaded on all factors and was dropped from the analysis. The factor *evidence availability* describes the responses of those who think it is important that evidence of an app's reliability and validity is available and are focused on the validity of their apps. This is distinct from *evidence accessibility*, which reflects the extent to which respondents could determine the reliability and validity of the apps they use. Further factors detailed *nonevidence use*, where participants did not consider validation or reliability important, and *resource*, which reflected responses indicating that price was more important; free apps were less likely to be used. We retained the participants' derived factor scores for each factor for analysis. These computed variables were all within the normal range of skewness, despite evidence availability (mean 4.16, SD 0.93; skew=−0.93) having a high average and nonevidenced use (mean 1.71, SD 0.66; skew=0.99) having a low average. Evidence accessibility (mean 3.39, SD

0.83; skew=−0.23) and resource (mean 2.44, SD 0.64; skew=0.17) were also within acceptable response ranges.

Table 1. The factor loadings of the exploratory factor analysis on the questionnaire items.

Questionnaire item	Evidence availability	Evidence accessibility	Nonevidenced use	Resource
It is important that there are reliability studies available for the apps I use	*0.91*[a]	−0.04	0.00	0.01
It is important that there are validity studies available for the apps I use	*0.90*	0.02	0.00	0.04
It is important that apps have validity	*0.80*	0.04	−0.06	−0.06
It is easy to determine the validity of the apps I use	0.09	*0.82*	0.10	0.03
It is easy to determine the reliability of the apps I use	0.08	*0.77*	0.20	0.01
It is difficult to determine the validity of the apps I use	0.12	*−0.71*	0.18	−0.03
It is difficult to determine the reliability of the apps I use	0.14	*−0.61*	0.33	−0.01
The apps I use in practice are easy to use	0.07	*0.38*	−0.16	−0.15
I do not deem the reliability of apps to be important	−0.07	0.11	*0.65*	−0.12
I do not consider the validity of apps to be important	−0.30	−0.02	*0.47*	0.14
The apps I use in practice are difficult to use	0.00	−0.27	*0.43*	0.23
Reliability of the apps is important	0.25	0.00	*−0.38*	0.13
The price is not important for me when selecting an app	0.01	0.14	−0.03	*0.56*
The price is important for me when selecting an app	−0.04	0.01	0.17	*−0.49*
I am more likely to use an app if it is free	−0.01	−0.15	0.21	*−0.44*
Other equipment is easier to use than apps for collecting the same data	−0.03	−0.12	0.35	*0.44*
Other equipment is harder to use than apps for collecting the same data	0.09	0.13	0.03	*−0.36*
I am more likely to use an app I have to pay for	0.26	0.14	0.22	0.27

[a]The text in italics highlights the strongest load of each item onto the factors.

Tests of the relationship between the categorical variables in this study (job role, level of education, sports type, level of athletes, coded data type, and perceived data type) were analyzed with chi-square tests of independence (using base R), with additional insight provided by the effect size Cramer *V* (using the R package *questionr*) [18]. Binomial linear models (using base R) were used to test the effect of scale variables (such as responses to the questionnaire) on binary outcomes (such as engagement with evidence-based apps or not).

Where the aforementioned categorical variables were used to test for a difference in conceptually and statistically similar dependent variables (the subscales of the questionnaire), multivariate analysis of variance (MANOVA) was used. The overall multivariate effects were tested with Pillai trace to be robust against violations of assumptions. MANOVA tests were conducted with base R, with the additional inference drawn from 95% CI of the omnibus test effect size, ω^2, using the *MOTE* package [19]. Where needed, post hoc follow-up tests on the MANOVA would involve analysis of variance and two-tailed Welch *t* tests for pairwise comparisons.

Results

Demographic Information

The survey received 335 responses. The mean age of the respondents was 32.9 (SD 9.9) years, with a range of 51 years. Respondents of 31 different nationalities completed the survey, with most of the survey responses received from the United Kingdom (107/335, 31.9%), Spain (107/335, 31.9%), and the United States (44/335, 13.1%). A total of 49 different sports were reported by the respondents. Table 2 provides an overview of respondents' demographic information.

Table 2. Respondents' demographic information (N=335).

Demographic	Respondent, n (%)
Sex	
Male	277 (82.7)
Female	55 (16.4)
Did not disclose	3 (0.9)
Level of education	
High school	3 (0.9)
Further education	18 (5.4)
Bachelors	89 (26.6)
Masters	184 (54.9)
Doctorate	41 (12.2)
Area of education	
Sport and exercise science	132 (39.4)
Strength and conditioning	65 (19.4)
Physical education	50 (14.9)
Sports coaching	23 (6.9)
Physiotherapy	13 (3.9)
Physical activity and health	11 (3.3)
Sports therapy	8 (2.4)
Nutrition	5 (1.5)
Psychology	3 (0.9)
Other	25 (7.5)
Current role	
Education	76 (22.7)
Applied sport science	148 (44.2)
Coaching	89 (26.6)
Other	22 (6.6)
Type of sports working with	
Team sports	116 (34.6)
Individual sports	84 (25.1)
Combination	93 (27.8)
Level of athletes working with	
Professional sport	124 (37.0)
Amateur sport	89 (26.6)
Combination	92 (27.5)

Smartphone and Tablet Use

Information on the general use of smartphones and tablets in sports practice is presented in Table 3. Respondents who answered *yes* to using smartphones and tablet devices were then asked to list apps used in their practice ("What apps do you currently use in your practice—please ensure you only refer to apps that do NOT require additional hardware."). Many respondents (Table 2) reported using either apps that required connection to additional external hardware (eg, GymAware) or apps with a primary function of logging training and activity data (eg, TeamBuildr). Of the 205 respondents who reported using direct data collection apps, the most frequent response (80/205, 39%) to the most important reason for using apps in practice was ease of use. Of the 205 respondents, 116 (56.5%) reported that the cost of apps was the least important reason for using apps. Among the 75 respondents who reported not using smartphones and tablet apps in their sports practice, the most frequently cited reason for not using them was a preference for other equipment (116/335, 34.6%).

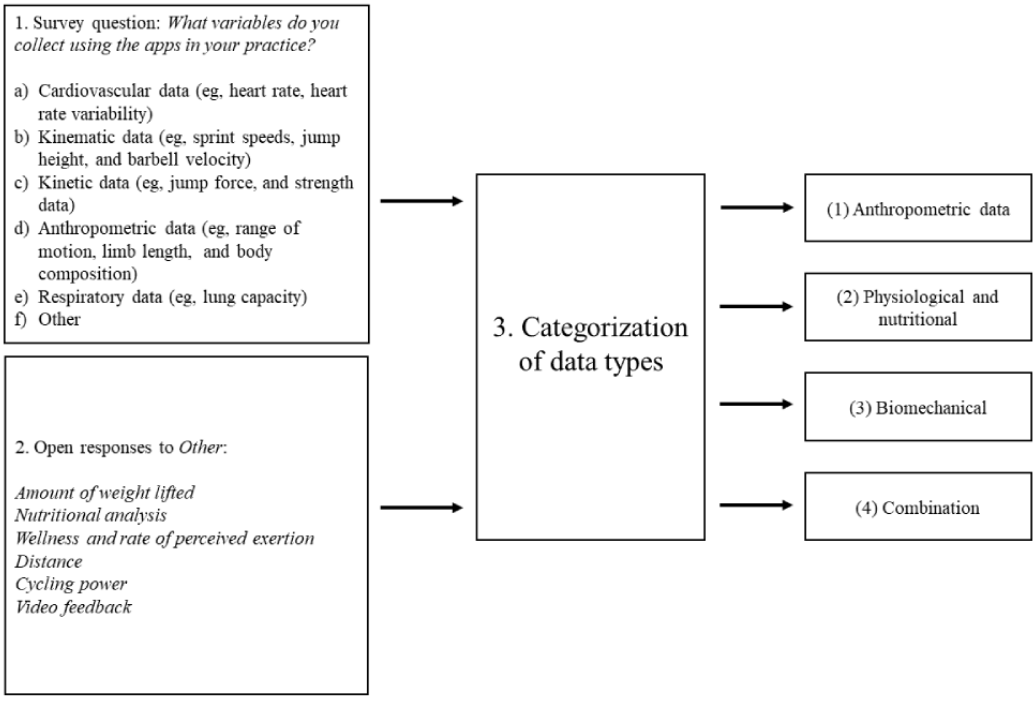
Table 3. Smartphone and tablet use (N=335).

Characteristics	Respondent, n (%)
Use smartphones or tablets in sports practice	
Yes	260 (77.6)
No	75 (22.4)
Type of app use in sports practice	
Only use apps with direct data collection	150 (44.8)
Use a combination of both direct and nondirect data collection apps	55 (16.4)
Only use apps with nondirect data collection (tracking apps)	40 (11.9)
Only use apps not compatible with a smartphone or tablet (ie, software that functions with PC)	15 (4.5)
Type of data collected using apps in sports practice	
Only collect anthropometric data	7 (2.1)
Only collect physiological and nutritional data	4 (1.2)
Only collect biomechanical data	153 (45.7)
Collecting a combination of data types	41 (12.2)

Types of Data Collection

Respondents who reported using a smartphone and/or tablet in their sports practice were asked to report what data they collected. We created 4 categories of data use based on these self-reports: (1) anthropometric (eg, joint range of motion, body composition, and limb length), (2) physiological and nutritional (eg, heart rate, heart rate variability, and dietary analysis), (3) biomechanical (eg, kinematic and kinetic data such as sprint speed, jump height, barbell velocity, and force data), and (4) a combination of the abovementioned 3 categories. Data types were coded from the reported apps to investigate how informed respondents were about the meaningful data that could be extracted from the apps (Figure 1). Interestingly, respondents' perceptions of the data they were collecting were significantly different from the data recorded by the apps (χ^2_4=230.9; P<.001; V=0.63). Respondents who were collecting a combination of data types (eg, anthropometric and biomechanical data) accurately reported collecting combined data (41/42, 98%). However, many who exclusively collected biomechanical data reported collecting a combination of data types (81/153, 52.9%) rather than solely biomechanical data (72/153, 47.1%). Furthermore, although many who were collecting no data reported not to collect data (58/75, 77%), other respondents within this group reported collecting biomechanical data (7/75, 9%) or a combination (10/75, 13%) of variables.

Figure 1. Categorization of data types from survey responses with examples of open-text responses to the "other" option.

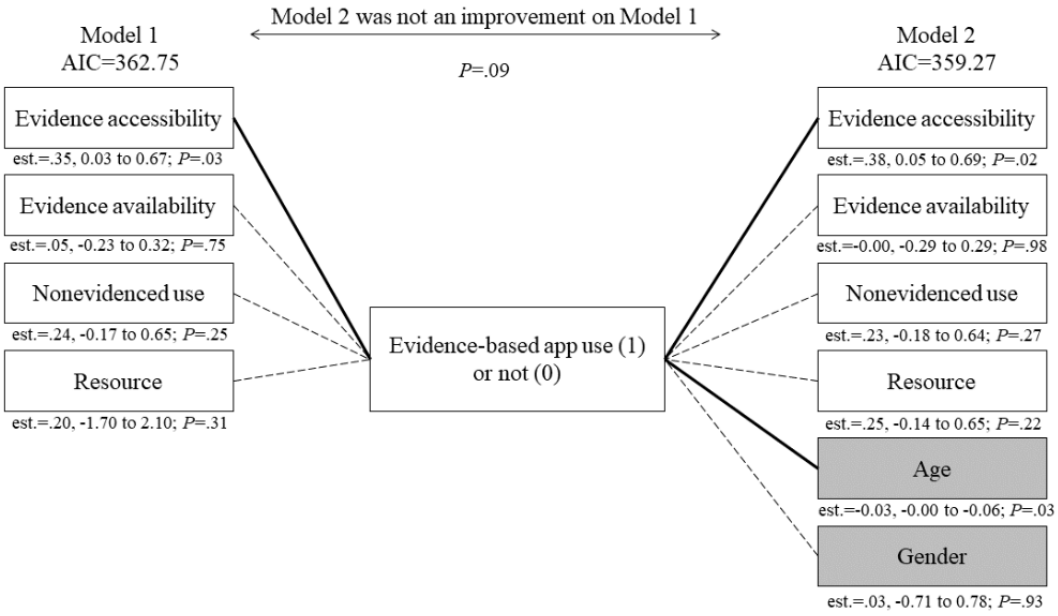


Evidence-Based App Use

Of the reported apps, 13 featured in studies that show no evidence of reliability and validity, whereas 15 apps appeared in studies that demonstrated reliability and validity. Only 1 app was validated but had no literature-demonstrated reliability, and 1 had reliability evidence but no validity evidence; therefore, these 2 apps were included in the 15 evidence-based apps. A total of 58.8% (197/335) respondents did not use any evidence-based apps. Fewer participants reported using 1 evidence-based app (95/335, 28.4%), 2 evidence-based apps (26/335, 7.7%), 3 evidence-based apps (12/335, 3.5%), and 4 evidence-based apps (2/335, 0.6%), and 3 respondents reported using 5 evidence-based apps; no one reported using more than 5 evidence-based apps. Given the limited variability in the number of apps used, we opted not to use the number of apps used as a variable for analysis. Rather, for more robust statistical analysis, the participants were dichotomized into *uses any evidenced-based apps or not*. This effectively presents the greatest behavioral distinction in our sample—engagement with apps or not.

A binomial general linear model using base R (R Foundation for Statistical Computing) was built to test the effect of the questionnaire factors on the use of evidence-based apps. The findings of the first model are summarized in Figure 2. Of the 4 factors (evidence availability, evidence accessibility, nonevidenced use, and resource), the only significant relationship with evidence-based app use was a higher score on evidence accessibility (β=.35, 95% CI 0.04-0.67; *P*=.03), that is, those who reported that it was easier to determine the validity and reliability of the apps were more likely to use those that had an evidence base. Respondents who found it more difficult to evaluate the evidence base of apps were less likely to use the evidence-based apps. None of the other questionnaire factors were significant (Figure 2).

Figure 2. A visual presentation of the results of the 2 linear binomial models. "Est." is the unstandardized β value predicting evidence-based app use (1) as opposed to not using evidence-based apps (0). Significant predictors are denoted by bold black lines, and nonsignificant predictors are denoted by dotted thin lines. Comparative model fits using the AIC are presented at the top, and the comparison of model fit tests by chi-square tests of variance is explained. AIC: Akaike Information Criteria.



A second model built using the same variables, with the addition of age and gender, was used to examine the general effect of respondent demographics and whether this affected evidence-based app use. This model did not show a significant improvement in explaining the variance in evidence-based app use (Figure 2). The summary of the second model again showed that evidence accessibility has a significant effect. Age had a significant negative effect on evidence-based app use (Figure 2), with younger adults more likely to use evidence-based apps.

Respondents reported information on their job role, education level, and types of athletes they worked with. As these were all discrete nominal variables with no numerical hierarchy—with more than 2 states—they were not included in the linear models. We tested the effect of athlete level (professional, amateur, combined, or no athletes) on engagement with evidence-based apps but found no differences (χ^2_3=1.3; *P*=.74; *V*=0.06). There were also no differences in the role of the respondent (education, applied sport science, coaching, or other; χ^2_3=5.2; *P*=.16; *V*=0.01), the type of athlete worked with (team sport, individual, combination, or nonathlete; χ^2_3=3.8; *P*=.28; *V*=0.01), or the level of education (bachelors, masters, or doctoral degree; χ^2_3=0.2; *P*=.89; *V*=0.03).

There was variability in engagement with evidence-based apps depending on the type of data being collected (χ^2_2=145.5; *P*<.001; *V*=0.67). This was explained primarily by the fact that those who were not collecting any data were predominantly using nonevidence-based apps (53/55, 96%) and those who were collecting combined data types preferred evidence-based

apps (40/41, 98%) to nonevidence-based apps. Participants solely collecting biomechanical data were split between those using nonevidence-based apps (66/153, 43.1%) and those using evidence-based apps (87/153, 56.9%). In general, those who were focused on collecting more complex data preferred to use apps with a clearer evidence base.

Questionnaire Responses by Demographics

Given that the questionnaire responses explained variance in engagement with evidence-based apps, it was of further interest to demonstrate any effect of professional activity and training on the questionnaire. There was no general effect of the type of athlete a respondent worked with and their responses to the questionnaire factors (Pillai trace=0.08; $F_{12,765}$=1.67; P=.07; ω^2=0.02, 95% CI 0.00-0.03). Similarly, there was no effect of level of athlete on responses to the questionnaire factors (Pillai trace=0.05; $F_{12,765}$=1.07; P=.39; ω^2=0.00, 95% CI 0.00-1.00), and there was no effect of respondents' level of education on responses to the questionnaire factors (Pillai trace=0.04; $F_{4,244}$=2.33; P=.06; ω^2=0.02, 95% CI 0.00-0.05).

Interestingly, the respondents' scores on the questionnaire factors varied by the type of data they were collecting (Pillai trace=0.11; $F_{8,488}$=3.51; P<.001; ω^2=0.06, 95% CI 0.01-0.09). This multivariate effect was explained univariate effects of data type on evidence accessibility ($F_{2,246}$=4.63; P=.01; ω^2=0.02, 95% CI 0.00-0.07) and evidence availability ($F_{2,246}$=7.22; P<.001; ω^2=0.04, 95% CI 0.00-0.09), but there was no effect of data type on nonevidenced use ($F_{2,246}$=1.24; P=.29; ω^2=0.00, 95% CI 0.00-0.02) or resource ($F_{2,246}$=1.72; P=.18; ω^2=0.00, 95% CI 0.00-0.03).

Participants collecting combined data types were those respondents with higher evidence accessibility scores (mean 3.76, SD 0.77), that is, they were those who indicated that it was easy to determine the evidence basis of the apps. This was demonstrated in subsequent Welch t tests, showing that those collecting combined data scored higher on evidence accessibility than those collecting biomechanical data (mean 3.33, SD 0.82; $t_{66.85}$=3.08; P=.003; Cohen d=0.44, 95% CI 0.19-0.89) or those collecting no data (mean 3.31, SD 0.87; $t_{91.49}$=3.08; P=.009; Cohen d=0.55, 95% CI 0.14-0.96). Those collecting biomechanical data and those collecting no data did not differ from each other ($t_{90.59}$=0.19; P=.85; Cohen d=0.03, 95% CI −0.34 to 0.28). The differences in evidence availability followed the same pattern. Respondents who considered reliability and validity more important were those collecting combined data (mean 4.60, SD 0.75). Welch t tests demonstrated that those collecting combined data scored higher on evidence availability than those collecting biomechanical data (mean 4.17, SD 0.90; $t_{73.89}$=3.13; P=.003; Cohen d=0.45, 95% CI 0.20-0.90) or those collecting no data (mean 3.90, SD 0.99; $t_{93.9}$=3.96; P<.001; Cohen d=0.82, 95% CI 0.39-1.24). Again, respondents collecting biomechanical data and respondents not collecting data did not differ from each other ($t_{88.18}$=1.78; P=.08; Cohen d=0.25, 95% CI −0.03 to 0.59).

Discussion

Principal Findings

This study aims to review the prevalence of mobile app use in the sport and exercise science industry, with a particular focus on apps with an evidence base to support their app for direct data collection. The main findings were that (1) 61.2% (205/335) of respondents reported using apps to directly collect data, (2) there was a misunderstanding in some users regarding the type of data being collected by the app, (3) biomechanical data were the most frequently collected type of data, (4) more than half of those using apps to collect data were doing so with apps that had no evidence base, and (5) perceived evidence availability and evidence accessibility had the strongest effects on evidence-based app use.

Consistent with the findings from the study by Jospe et al [20], the most frequently reported reason for using smartphone apps was a perceived ease of use, which is consistent with the broader literature on mobile technology [21,22]. However, the most frequently cited barrier to not using smartphones and tablets in practice was a preference for other equipment. Furthermore, one-fifth of the respondents reported not using apps because of a lack of compatibility with their current resources. This has previously been highlighted by Ravenek and Alvarez [23], whereby practitioners may not have the appropriate infrastructure to support mobile device use (eg, internet connectivity). Consequently, those who do not use apps in their sports practice may be hindered by structural and operational constraints specific to their respective workplaces rather than a lack of motivation to engage with smartphones. Conversely, using smartphones within health and other related contexts is still considered a new concept [23,24], and some practitioners may be skeptical of new technology, particularly if they perceive a lack of knowledge in using smartphone apps. Evidence of this concern is consistent in the broader literature, as practitioners are uncomfortable in both prescribing apps to clients and patients and using them in personal practice if they feel they do not possess the appropriate prerequisite knowledge [20,24,25]. Another consistent theme in the existing literature is greater smartphone use by younger respondents [26]. In this study, we found that age had a significant negative effect on evidence-based app use, whereby younger adults were more likely to use evidence-based apps. This investigation has predominantly focused on how mobile technologies are used in sport and exercise practices, as opposed to why the technology is used. Further investigations are therefore required to examine the possible reasons for not using smartphone and tablet technology.

To our knowledge, this is the first investigation to explore the prevalence of smartphone and tablet use for direct data collection, that is, no additional hardware in conjunction with app use. Respondents were asked, "What apps do you currently use in your practice—please ensure you only refer to apps that do NOT require additional hardware?". A total of 55 respondents did not use apps that used a smartphone or tablet as a direct data collection tool, that is, they did not use the internal hardware of a smartphone or tablet.

Typical responses were those reporting the use of an app that stores information entered by a user. Examples of this include logging repetitions, sets, and loads used in resistance training sessions or recording heart rate determined by an external device typically paired to a smartphone via Bluetooth. When these types of users were removed, 61.2% (205/335) of our sample were using a smartphone or tablet to directly collect data. There were further misunderstandings when respondents were asked to report what types of data they collected (eg, anthropometric, physiological, and biomechanical). For example, some respondents reported collecting both biomechanical and anthropometric data but only listed apps designed to measure biomechanical data, such as jump height and sprint speed. The first explanation for this misunderstanding is that the survey was provided to all respondents in British English only, despite having responses from 31 different countries. Of those that erroneously reported the types of data they were collecting from their respective apps, 51% (50/98) were from a country that did not have English as an official language. Questions can be interpreted differently depending on the language it is asked in [27]. It may have been difficult to understand the difference between recording inputted data and directly collecting data with a smartphone app and the difference in data types. As many of these respondents were from English-speaking countries, a second explanation is a general misunderstanding of which apps they use and their capabilities. Bromilow et al [14] found that 56% of their respondents could not identify which smartphone apps they used.

Despite some misunderstanding of the types of data being collected, the number of respondents reporting the use of a smartphone or tablet was much higher than that reported by Bromilow et al [14], who found that only 9% of their sample used a smartphone for direct data collection. Biomechanical data, such as kinetics related to vertical jump performance, were the most frequent (153/335, 45.7%) type of data collected. The results of our survey suggest that this is because of the perceived availability of peer-reviewed literature. Evidence accessibility was the only significant element of the model for evidence-based app use (β=.35, P=.03), and respondents were more likely to use apps with an evidence base if they perceived it was easier to find evidence of validity and reliability. This is unsurprising considering the current wealth of literature that has focused on the validation of apps used to collect various biomechanical variables. MyJump2 is a smartphone app that has been shown to validly and reliably estimate vertical jump performance in multiple populations [28,29]. The app has featured so significantly in the peer-reviewed literature that a narrative review has been provided by Sharpe et al [30]. There is significant cost and expertise required for collecting these type of data using more traditional methods such as a force plate [10], which may explain why cost-effective and user-friendly apps—investigations of their respective validity and reliability—are popular in this particular discipline.

In contrast, although there is some evidence demonstrating valid and reliable cardiovascular measures, such as heart rate variability [31], Muntaner-Mas et al [32] suggest that there is a general lack of peer-reviewed literature on apps related to cardiorespiratory fitness. This is in line with our findings that only 7% of respondents reported using apps to collect physiological data variables. Peart et al [10] suggested that there is now a stable body of research on apps that collect biomechanical data. This seems promising; however, of the respondents stating that they only collected biomechanical data, 43.1% (66/153) did not use evidence-based apps. Therefore, although this area has a number of apps supported by the literature, there is also more choice available and an increased risk of selecting nonvalidated apps. For example, smartphone apps that collect kinematic and kinetic data of barbell exercises continue to be developed and made commercially available for validity and reliability studies. It is, therefore, possible that some users download an app based on popularity rather than their quality, with regard to validity and reliability, as a result of *app overload* [24].

In total, 59.0% (121/205) of the respondents did not use evidence-based apps, that is, where peer-reviewed literature has provided evidence of acceptable reliability and validity. The literature assessing the validity and reliability of mobile device hardware used in other contexts is extremely limited, making it difficult to draw direct comparisons with other contexts. Many apps are commercially available to promote behavior change, such as smoking cessation [33], weight loss [34], and suicide prevention [35]. Haskins et al [33] identified 6 smoking cessation apps with some level of scientific support, of which only 2 featured in any top 50 app lists in web-based app stores. The authors concluded that scientifically informed apps were underutilized. Although not directly comparable, our findings are consistent with other app contexts, showing that many users adopt software apps with no scientific support. There is consensus in the broader literature that there are challenges in highlighting the availability of scientifically informed apps to a user base, and, as demonstrated by our findings, the sport and exercise science community is not an exception to this. The use of evidence-based apps was significantly explained by the *evidence accessibility* factor in our model. This factor reflects the extent to which respondents could determine the reliability and validity of the apps they use, independent of the amount of evidence available. Therefore, even if users had access to scientific information, their self-reported ability to understand evidence was the main driver of their choice to use evidence-based apps. In practical terms, this means that although the existence of evidence is important, whether this evidence is effectively communicated to the consumer is of higher importance. Interestingly, the volume of data collected by apps influenced the likelihood of choosing evidence-based apps. Respondents collecting a combination of data types considered reliability and validity more important when selecting data. We speculate that those who were collecting multiple performance variables would have to use multiple apps and were therefore more aware of data collection apps currently available on the market. Such individuals are potentially more familiar with which apps have an existing evidence base in the literature. This is consistent with the existing literature, which indicates that smartphone app proficiency is more closely related to individual interest rather than the level of education [14]. Our results demonstrated that the level of education did not significantly affect the adoption of evidence-based apps.

Limitations

The limitations of this study must be addressed. First, a limitation of survey research, in general, is the risk that they are more likely to be completed by people who have a preconceived idea about the topic in question, that is, people who use apps may be more likely to be interested in taking part. For example, Bromilow et al [14] found that 99% of their sample reported using smartphone apps in their sport and exercise science practices. In this study, we found that smartphone and tablet use was less prevalent than previously reported [14], with 78% of respondents reporting the use of a smartphone or tablet in their practice. Both this investigation and the previous investigation by Bromilow et al [14] may possess a nonresponse bias [36], which overestimates smartphone and tablet use in sport- and exercise-related practices. In a related context, Jospe et al [20] suggested a nonresponse bias as a limitation of their investigation of sport dieticians' use of smartphone apps.

Apps that had their reliability and validity findings reported in the academic literature were considered evidence-based apps for the purpose of this investigation. However, it is possible that the software developers have conducted internal validation testing and calibration. As peer review and publication of academic literature can take a substantial amount of time, it is plausible that some mobile software apps are valid and reliable but have not yet been reported and published in the literature. In relation to this, another potential limitation is that some respondents may not have known the difference between validity and reliability. However, we did not find any differences between the reliability and validity questions in our survey. In addition, we did not ask participants how they may have used reviews and user ratings to inform their app selection. This was beyond the scope of this study, and further research is required to qualitatively investigate how users decide to select the apps they use in practice.

This investigation provides insight into the broad use of mobile technologies to directly collect data in sport-related and exercise science–related fields. The results of this study demonstrate that practitioners show a greater preference for using smartphones and tablet devices to collect biomechanical data such as sprint velocity and jump performance variables. This may be because of a greater prevalence of peer-reviewed literature, which has provided evidence of valid and reliable apps, and because practitioners can access this information. When practitioners perceive that it is easier to determine the quality of apps, this leads to increased adoption of evidence-based apps. Therefore, there are 2 key implications for app developers. First, app developers should seek independent research to validate their apps. Second, app developers must consider how they market their products. Using journal articles to select apps is ineffective [14], and app developers should look to provide clear signposting to the scientific support of their software in alternative ways, such as app store descriptions and social media.

Conflicts of Interest

None declared.

References

1. Romero-Franco N, Jiménez-Reyes P, Castaño-Zambudio A, Capelo-Ramírez F, Rodríguez-Juan JJ, González-Hernández J, et al. Sprint performance and mechanical outputs computed with an iPhone app: comparison with existing reference methods. Eur J Sport Sci 2017 May;17(4):386-392. [doi: 10.1080/17461391.2016.1249031] [Medline: 27806673]
2. Balsalobre-Fernández C, Cardiel-García M, Jiménez SL. Bilateral and unilateral load-velocity profiling in a machine-based, single-joint, lower body exercise. PLoS One 2019;14(9):e0222632 [FREE Full text] [doi: 10.1371/journal.pone.0222632] [Medline: 31525245]
3. Balsalobre-Fernández C, Marchante D, Baz-Valle E, Alonso-Molero I, Jiménez SL, Muñóz-López M. Analysis of wearable and smartphone-based technologies for the measurement of barbell velocity in different resistance training exercises. Front Physiol 2017;8:649 [FREE Full text] [doi: 10.3389/fphys.2017.00649] [Medline: 28894425]
4. Haynes T, Bishop C, Antrobus M, Brazier J. The validity and reliability of the My Jump 2 app for measuring the reactive strength index and drop jump performance. J Sports Med Phys Fitness 2019 Feb;59(2). [doi: 10.23736/s0022-4707.18.08195-1]
5. Cox RW, Martinez R, Baker R, Warren L. Validity of a smartphone application for measuring ankle plantar flexion. J Sport Rehabil 2018 May 01;27(3):1-3. [doi: 10.1123/jsr.2017-0143] [Medline: 29140194]
6. Furness J, Schram B, Cox A, Anderson SL, Keogh J. Reliability and concurrent validity of the iPhone Compass application to measure thoracic rotation range of motion (ROM) in healthy participants. PeerJ 2018;6:e4431 [FREE Full text] [doi: 10.7717/peerj.4431] [Medline: 29568701]
7. Lee JW, Cai M, Yung P, Chan KM. Reliability, validity, and sensitivity of a novel smartphone-based eccentric hamstring strength test in professional football players. Int J Sports Physiol Perform 2018 May 01;13(5):620-624. [doi: 10.1123/ijspp.2017-0336] [Medline: 29283704]
8. Coppetti T, Brauchlin A, Müggler S, Attinger-Toller A, Templin C, Schönrath F, et al. Accuracy of smartphone apps for heart rate measurement. Eur J Prev Cardiol 2017 Aug;24(12):1287-1293. [doi: 10.1177/2047487317702044] [Medline: 28464700]

9. Peart DJ, Shaw MP, Rowley CG. An investigation into a contactless photoplethysmographic mobile application to record heart rate post-exercise: implications for field testing. Biomed Hum Kinet; 7. Epub ahead of print 11 August 2015:95-99. [doi: 10.1515/bhk-2015-0015]
10. Peart DJ, Balsalobre-Fernández C, Shaw M. Use of mobile applications to collect data in sport, health, and exercise science: a narrative review. J Strength Cond Res 2019 Apr;33(4):1167-1177. [doi: 10.1519/JSC.0000000000002344] [Medline: 29176384]
11. Radtke KL, Porcari JP, Marx R, Mikat R, Doberstein S, Ryskey A, et al. Ability of the LeanScreen app to accurately assess body composition. Med Sci Sports Exerc 2018;50:159. [doi: 10.1249/01.mss.0000535610.04677.17]
12. Shaw MP, Robinson J, Peart DJ. Comparison of a mobile application to estimate percentage body fat to other non-laboratory based measurements. Biomed Hum Kinet 2017;9:98. [doi: 10.1515/bhk-2017-0014]
13. Wagner DR, Castañeda F, Bohman B, Sterr W. Comparison of a 2D iPad application and 3D body scanner to air displacement plethysmography for measurement of body fat percentage. J Hum Nutr Diet 2019 Dec;32(6):781-788. [doi: 10.1111/jhn.12687] [Medline: 31334579]
14. Bromilow L, Stanton R, Humphries B. A structured e-investigation into the prevalence and acceptance of smartphone applications by exercise professionals. J Strength Cond Res 2020 May;34(5):1330-1339. [doi: 10.1519/JSC.0000000000003301] [Medline: 32028462]
15. Molina MD, Sundar SS. Can mobile apps motivate fitness tracking? A study of technological affordances and workout behaviors. Health Commun 2020 Jan;35(1):65-74. [doi: 10.1080/10410236.2018.1536961] [Medline: 30358424]
16. Leicester. Ethics guidelines for internet-mediated research. British Psychological Society. 2017. URL: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf [accessed 2021-04-19]
17. Selker R, Love J, Dropmann D. jmv: the 'jamovi' analyses. URL: https://cran.r-project.org/web/packages/jmv/index.html [accessed 2021-04-19]
18. Barnier J, Briatte F, Larmarange J. questionr: functions to make surveys processing easier. URL: https://cran.r-project.org/web/packages/questionr/index.html [accessed 2021-04-19]
19. Buchanan EM. MOTE. RDocumentation. URL: https://www.rdocumentation.org/packages/MOTE/versions/1.0.2 [accessed 2021-04-19]
20. Jospe MR, Fairbairn KA, Green P, Perry TL. Diet app use by sports dietitians: a survey in five countries. JMIR Mhealth Uhealth 2015 Jan 22;3(1):e7 [FREE Full text] [doi: 10.2196/mhealth.3345] [Medline: 25616274]
21. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 06;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
22. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mhealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
23. Ravenek M, Alvarez L. Use of mobile ‘apps’ in occupational therapy: therapist, client and app considerations to guide decision-making. World Fed Occup Ther Bull 2016 Mar 31;75(1):43-49. [doi: 10.1080/14473828.2016.1162430]
24. Rowe M, Sauls B. The use of smartphone apps in clinical practice: a survey of South African physiotherapists. S Afr J Physiother 2020;76(1):1327 [FREE Full text] [doi: 10.4102/sajp.v76i1.1327] [Medline: 32391442]
25. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Aug;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
26. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Info Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
27. Zavala-Rojas D, Saris WE. Measurement invariance in multilingual survey research: the role of the language of the questionnaire. Soc Indic Res 2017 Nov 14;140(2):485-510. [doi: 10.1007/s11205-017-1787-x]
28. Cruvinel-Cabral RM, Oliveira-Silva I, Medeiros A, Claudino JG, Jiménez-Reyes P, Boullosa DA. The validity and reliability of the "My Jump App" for measuring jump height of the elderly. PeerJ 2018;6:e5804 [FREE Full text] [doi: 10.7717/peerj.5804] [Medline: 30356977]
29. Gallardo-Fuentes F, Gallardo-Fuentes J, Ramírez-Campillo R, Balsalobre-Fernández C, Martínez C, Caniuqueo A, et al. Intersession and intrasession reliability and validity of the My Jump App for measuring different jump actions in trained male and female athletes. J Strength Cond Res 2016 Jul;30(7):2049-2056. [doi: 10.1519/JSC.0000000000001304] [Medline: 27328276]
30. Sharp AP, Cronin JB, Neville J. Using smartphones for jump diagnostics: a brief review of the validity and reliability of the My Jump App. Strength Cond J 2019;41:96-107. [doi: 10.1519/SSC.0000000000000472]
31. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography. Int J Sports Physiol Perform 2017 Nov 01;12(10):1324-1328. [doi: 10.1123/ijspp.2016-0668] [Medline: 28290720]

32. Muntaner-Mas A, Martinez-Nicolas A, Lavie CJ, Blair SN, Ross R, Arena R, et al. A systematic review of fitness apps and their potential clinical and sports utility for objective and remote assessment of cardiorespiratory fitness. Sports Med 2019 Apr;49(4):587-600 [FREE Full text] [doi: 10.1007/s40279-019-01084-y] [Medline: 30825094]
33. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
34. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
35. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
36. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):1-9 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]

Abbreviations

EFA: exploratory factor analysis
MANOVA: multivariate analysis of variance

Edited by L Buis; submitted 25.06.20; peer-reviewed by M Altini, P Lee, L Becker; comments to author 14.09.20; revised version received 01.10.20; accepted 22.03.21; published 13.05.21.

Please cite as:
Shaw MP, Satchell LP, Thompson S, Harper ET, Balsalobre-Fernández C, Peart DJ
Smartphone and Tablet Software Apps to Collect Data in Sport and Exercise Settings: Cross-sectional International Survey
JMIR Mhealth Uhealth 2021;9(5):e21763
URL: https://mhealth.jmir.org/2021/5/e21763
doi:10.2196/21763
PMID:33983122

Original Paper

Engagement With a Health Information Technology–Augmented Self-Management Support Program in a Population With Limited English Proficiency: Observational Study

Leah Machen[1], MD; Margaret A Handley[1], MPH, PhD; Neil Powe[1], MD, MPH, MBA; Delphine Tuot[1], MDCM, MAS

University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Delphine Tuot, MDCM, MAS
University of California, San Francisco
1001 Potrero Ave
San Francisco, CA, 94110
United States
Phone: 1 415 206 3784
Email: Delphine.tuot@ucsf.edu

Abstract

Background: Limited English proficiency (LEP) is an important driver of health disparities. Many successful patient-level interventions to prevent chronic disease progression and complications have used automated telephone self-management support, which relies on patient activation and communication to achieve improved health outcomes. It is not clear whether these interventions are similarly applicable to patients with LEP compared to patients with English proficiency.

Objective: The objectives of this study were as follows: (1) To examine the impact of LEP on patient engagement (primary outcome) with a 12-month language-concordant self-management program that included automated telephone self-management support, designed for patients with chronic kidney disease (CKD). (2) To assess the impact of LEP on change in systolic blood pressure (SBP) and albuminuria (secondary outcomes) resulting from the self-management program.

Methods: This was a secondary analysis of the Kidney Awareness Registry and Education (KARE) pilot trial (NCT01530958) which was funded by the National Institutes of Health in August 2011, approved by the University of California Institutional Review Board in October 2011 (No. 11-07399), and executed between 2013 and 2015. Multivariable logistic and linear models were used to examine various facets of patient engagement with the CKD self-management support program by LEP status. Patient engagement was defined by patient's use of educational materials, completion of a health coaching action plan, and degree of participation with automated telephone self-management support. Changes in SBP and albuminuria at 12 months by LEP status were determined using multivariable linear mixed models.

Results: Of 137 study participants, 53 (38.7%) reported LEP, of which 45 (85%) were Spanish speaking and 8 (15%) Cantonese speaking. While patients with LEP and English proficiency similarly used the program's educational materials (85% [17/20] vs 88% [30/34], *P*=.69) and completed an action plan (81% [22/27] vs 74% [35/47], *P*=.49), those with LEP engaged more with the automated telephone self-management support component. Average call completion was 66% among patients with LEP compared with 57% among those with English proficiency; patients with LEP requested more health coach telephone calls (*P*=.08) and had a significantly longer average automated call duration (3.3 [SD 1.4] min vs 2.2 [1.1 min], *P*<.001), indicating higher patient engagement. Patients with LEP randomized to self-management support had a larger, though nonstatistically significant (*P*=.74), change in SBP (–4.5 mmHg; 95% CI –9.4 to 0.3) and albuminuria (–72.4 mg/dL; 95% CI –208.9 to 64.1) compared with patients with English proficiency randomized to self-management support (–2.1 mmHg; 95% CI –8.6 to 4.3 and –11.1 mg/dL; 95% CI –166.9 to 144.7).

Conclusions: Patients with LEP with CKD were equally or more engaged with a language-concordant, culturally appropriate telehealth intervention compared with their English-speaking counterparts. Augmented telehealth may be useful in mitigating communication barriers among patients with LEP.

Trial Registration: ClinicalTrials.gov NCT01530958; https://clinicaltrials.gov/ct2/show/NCT01530958

(JMIR Mhealth Uhealth 2021;9(5):e24520) doi:10.2196/24520

KEYWORDS

automated telehealth intervention; limited English proficiency; mHealth; language problems; telehealth; SMS; text message

Introduction

In 2019 the US Census Bureau reported that there were over 25 million people living in the United States classified as having limited English proficiency (LEP), defined by not answering "very well" regarding their ability to speak English [1]. Not only is LEP linked to lower socioeconomic status and poverty, but it is also a known predictor of poorer health status, decreased preventive care, and less access to medical care [2,3]. The obstacles to receiving medical care for these patients are further compounded by communication barriers, facilitating poor engagement with health care providers, worse interpersonal care and patient satisfaction, and less patient education [4-6]. Because of these factors, the LEP population may be a growing marginalized group of patients with high disease burden and poor connections with the health care system [1].

Automated telephone self-management support systems are a form of health information technology (HIT) that use patient vignettes to deliver education and self-management tools to patients and interactive voice response with an option to request clinician phone calls to engage patients in their health care. Automated telephone self-management support systems have been shown to improve clinical outcomes such as systolic blood pressure (SBP), depressive symptoms, and obesity, and reduce hospital admissions and mortality for patients with heart failure [6-10]. They have also been shown to have high levels of patient engagement, even among older adults who may be unable to use other forms of HIT due to limited vision, literacy, or technology knowledge [11]. This dynamic and interactive tool offers a unique opportunity to reach marginalized populations with poor health access and promote engagement in populations such as those with LEP.

Despite long-time awareness of the existence of this population and the gaps in their care, providers are only recently beginning to develop targeted health interventions to improve access and outcomes for patients with LEP. Patient engagement, encompassing patient activation which is defined as having the motivation, knowledge, skills, and confidence to make effective decisions to manage one's health, has been shown to be associated with improved patient lifestyle choices, adherence, and chronic disease management [12-18]. Thus, finding novel ways to augment patient engagement will serve as one of many essential pathways to improve health care access and outcomes among the LEP population.

Automated telephone self-management support systems represent a group of novel interventions that can be tailored to specific patient populations. We sought to explore whether the LEP status impacted patient engagement with a language-concordant self-management program that featured automated telephone self-management support systems among participants with chronic kidney disease (CKD) randomized to the intervention arm of the Kidney Awareness Registry and Education (KARE) pilot trial. We hypothesized that patients with LEP would have higher levels of engagement than those with English proficiency given the novelty of a language-concordant intervention in an environment with a paucity of non-English self-management support materials.

Methods

Study Design

We conducted a retrospective analysis of the KARE pilot trial to assess participant engagement with and impact of an HIT-augmented comprehensive self-management support program on change in SBP and albuminuria among patients with limited (vs adequate) English proficiency. Details of the KARE study have been previously described [19]. In brief, KARE was a 2 × 2 factorial pilot randomized controlled trial that took place in 2 primary care clinics in San Francisco's public health care delivery system. The study was funded by the National Institutes of Health in August 2011 and approved by the University of California Institutional Review Board in October 2011 (No. 11-07399). A total of 137 patients were enrolled in the trial, which was executed between 2013 and 2015. Overall results of the pilot trial have been published previously [20].

The KARE study had 2 levels of randomization. First, within each clinic, primary care practice teams consisting of several physicians (including trainees), 1 nurse, nurse practitioners, medical assistants, and behaviorists, were randomized 1:1 to 1 of 2 arms with a random number generator: access to a CKD registry versus usual care registry. Second, within 6 months of the provider-level randomization, eligible patients were recruited to a baseline visit and were randomized within each provider 1:1 to participate in a year-long comprehensive CKD self-management program.

Eligible patients included adults (≥18 years) with CKD, defined by 2 values of estimated glomerular filtration rate of 15-60 mL/min/1.73 m^2 or albuminuria (urine dipstick ≥1+ or urine albumin-to-creatinine ratio >30 mg/g) documented in the electronic health record on 2 occasions at least 90 days apart, who had contact with their primary health care team at least once within the past 2 years and spoke English, Spanish, or Cantonese. Patients who spoke Spanish or Cantonese were monolingual and did not speak English. Patients were excluded from the study if they were recipients for kidney transplantation, pregnant, or were unlikely to benefit from the self-management support program due to hearing or visual impairment, impaired cognition or severe mental illness, or a life expectancy less than 6 months.

Intervention

The KARE interventions have been previously described in detail and found to be acceptable among providers and patients [19,21]. In brief, the provider intervention consisted of an electronic health record–enabled CKD registry tool with "in-reach" and "outreach" elements to support team-based management of CKD. The patient intervention was a comprehensive CKD self-management support program with

3 distinct elements. The first element consisted of language-concordant, low-literacy written patient educational materials [22] couriered to patients at months 1, 4, and 8. The second element was a language-concordant and culturally tailored interactive automated telephone self-management program with 26 modules delivered every other week that provided education and self-management strategies related to topics pertinent to kidney health: basics of kidney disease and its association with hypertension; importance of participation in healthy behaviors (diet, physical activity, smoking cessation, stress reduction); avoidance of nonsteroidal anti-inflammatory medications; participation and preparation for clinic visits; complementary medication use; medication adherence; and glycemic control. In addition to providing educational content, the automated phone calls included "quiz questions" which patients would answer with a touch-tone phone throughout the module. These questions promoted engagement with the program and allowed patients to hear additional content or request telephone calls from their health coach. The third element of the CKD self-management support program consisted of telephone-based health coaching delivered by lay bilingual health coaches trained in motivational interviewing and action planning who called patients upon their request via the interactive telephone program and on an as-needed basis. Multimedia Appendix 1 includes an example English script for 1 week's content related to use of nonsteroidal anti-inflammatory medications, as well as the corresponding health coach guide to promote the development of an action plan related to this topic.

Study Outcomes

The primary outcome of this secondary analysis was patient engagement with the intervention determined a priori by the following measures which have been previously defined as engagement metrics for similar interventions [23]: percentage of users who engaged with any aspect of the automated system (defined by responding to at least one automated telephone self-management support call), percentage of automated calls completed, percentage of users who completed at least 80% of calls, and percentage of patients who created at least one action plan. We also a priori identified other engagement metrics that we believed fit the CKD self-management support program, including self-reported use of educational materials during study visits, duration of participation in the automated telephone calls determined by the telephone software (longer duration suggesting greater engagement due to interactive answering of health questions, repeating patient vignettes, and call completion), and number of requests for health coach callbacks from study participants during their automated telephone call (greater number of requests suggesting greater engagement).

Secondary outcomes of this study included changes in SBP and albuminuria severity (spot urinary albumin-to-creatinine ratio) from baseline to 12 months, both ascertained at study visits. SBP was measured with a digital blood pressure monitor (model HEM-907X; Omron), using the average of 3 SBP measurements in the right arm after the patient sat quietly for 5 min.

Covariates

At the baseline visit, patient sociodemographic data (age, gender, race/ethnicity, education, income, insurance status) were self-reported, as were comorbidity data (diabetes, coronary artery disease, hyperlipidemia). Food insecurity and health literacy were ascertained using validated screening questionnaires [24].

Statistical Analysis

We examined baseline characteristics of KARE participants and tested for differences by LEP status using chi-square tests for categorical variables and nonparametric Kruskal–Wallis tests for continuous variables. Multivariable logistic and linear regression models were used to determine differences in engagement with the self-management support program by LEP status adjusted for age, sex, and variables that differed by LEP status (race/ethnicity, education, insurance status). Impact of the self-management support program on change in SBP at 12 months and change in albuminuria by LEP status were determined using linear mixed models accounting for clustering by provider and controlling for sociodemographic variables and baseline measures of SBP and albuminuria, respectively.

Results

Participant Demographics

Of the 137 KARE participants, roughly half were male (66/137, 48.2%) with a mean age of 55.3 (SD 12.2) years; 8 of 137 participants were White (5.8%), 59 Black (43.1%), 49 Hispanic (35.8%), and 20 Asian (14.6%); 1 (0.7%) participant declined to answer. Over one-third (53/137, 38.7%) of participants reported LEP with primary languages of Spanish (45/53, 85%) and Cantonese (8/53, 15%); 71 out of 137 participants were high-school educated (51.8%) and 47/137 college (34.3%) educated, with lower numbers in the LEP group, 27/53 (51%) and 7/53 (13%), respectively, compared with the English-speaking group (P<.001 for both). In the overall cohort, 34/137 (24.8%) participants were uninsured or covered by a health care access program to subsidize medical care for uninsured residents of San Francisco, with 21/53 (40%) participants being in the LEP group and 13/84 (15%) in the English-speaking group (P<.001). Over three-quarters (91/121, 75.2%) of participants who provided these data reported annual income less than US $15,000 and 72 participants (52.6%) reported food insecurity, without any differences by LEP status. All study participants had CKD, of which 38.7% (53/137) had hypertension and 51.8% (71/137) had diabetes (Table 1).

Table 1. Sociodemographic characteristics of the study patients.

Characteristics	Overall (n=137)	Non-LEP (n=84)	LEP (n=53)	*P* value
Age (years), mean (SD)	55.3 (12.2)	56.3 (10.9)	53.7 (13.9)	.22
Sex, n (%)				.11
Male	66 (48.2)	45 (53.6)	21 (39.62)	
Female	71 (51.8)	39 (61.31)	32 (60.4)	
Race/Ethnicity, n (%)				<.001
White	8 (5.8)	8 (9.6)	0 (0.0)	
Black	59 (43.1)	59 (71.1)	0 (0.0)	
Hispanic	49 (35.8)	4 (4.8)	45 (84.9)	
Asian	20 (14.6)	12 (14.5)	8 (15.1)	
Language, n (%)				<.001
English	84 (61.3)	84 (100.0)	0 (0.0)	
Spanish	45 (32.8)	0 (0.0)	45 (84.9)	
Cantonese	8 (5.8)	0 (0.0)	8 (15.1)	
Education, n (%)				<.001
Primary school	19 (13.9)	0 (0.0)	19 (35.9)	
High school/Technical education	71 (51.8)	44 (52.4)	27 (50.1)	
College	47 (34.3)	40 (47.6)	7 (13.2)	
Insurance, n (%)				<.001
None or HSF	34 (24.8)	13 (15.5)	21 (39.6)	
Medicaid	61 (44.5)	44 (52.4)	17 (32.1)	
Medicare	37 (27.0)	26 (30.9)	11 (20.8)	
Other	5 (3.6)	1(1.2)	4 (7.6)	
Income[a], n (%)				.43
<15K	91 (75.2)	56 (77.8)	35 (71.4)	
15-50K	30 (24.8)	16 (22.2)	14 (28.5)	
Food insecurity	72 (52.6)	42 (58.3)	30 (56.6)	.45
Health literate	101 (73.7)	63 (75.0)	38 (71.7)	.67
Hypertension (≥140/90 mmHg)	53 (38.7)	37 (69.8)	16 (30.2)	.11
Diabetes	71 (51.8)	39 (54.9)	32 (45.1)	.11
CKD stage, n (%)				.23
CKD stages 1 and 2	46 (33.6)	25 (29.8)	21 (39.6)	
CKD Stages 3 and 4	91 (66.4)	59 (70.2)	32 (60.4)	

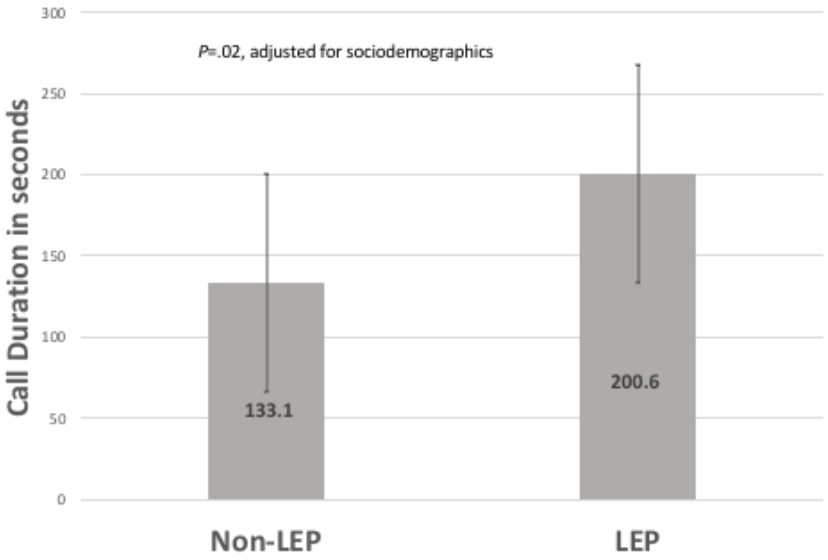
[a]n=137 for all rows except income, for which n=121 (n=72 for non-LEP and n=49 for LEP).

Primary Outcomes

Participant engagement measures pertinent to all 3 components of the CKD self-management support program included self-reported use of written education materials, development of at least one action plan, and degree of interaction with the automated telephone self-management support program. The overall use of patient education materials was 87% (47/54) and 57/74 (77%) participants developed an action plan. There were similar rates of self-reported use of educational materials among participants in the non-LEP and LEP group: 88% (30/34) and 85% (17/20), respectively (*P*=.73). Similarly, 35/47 (74%) participants in the non-LEP group and 22/27 (81%) participants in the LEP group developed at least one health-oriented action plan (*P*=.49). By contrast, engagement with the automated telephone self-management support differed by LEP status. While nearly all participants completed at least one automated telephone self-management support module (only 1 patient with English proficiency did not complete any automated telephone self-management support modules), the average completion rate of all modules was 57% among those with English proficiency and 66% among those with LEP. In multivariable

linear regression, LEP status was positively associated with a higher automated telephone self-management support module completion rate (coefficient=0.3/10% higher completion), although this was not statistically significant (*P*=.75). "High engagers" were classified as participants who completed at least 80% of automated phone calls, of which 61% (17/28) were in the non-LEP group and 39% (11/28) were in the LEP group; however, there was a nonstatistically significant difference in this regard (*P*=.70). Among participants who did complete the automated telephone self-management support calls, the mean number of health coach callbacks requested was significantly larger in the LEP group compared with the non-LEP group (16 [SD 14.8] vs 11 [SD 10.6]; *P*=.004); however, in the adjusted model this outcome did not reach statistical significance (*P*=.08). The average call duration among the LEP group was 200.6 seconds compared with 133.1 seconds in the non-LEP group, a difference which was statistically significant when adjusted for age, sex, race/ethnicity, education, and insurance status (*P*=.02; Figure 1)

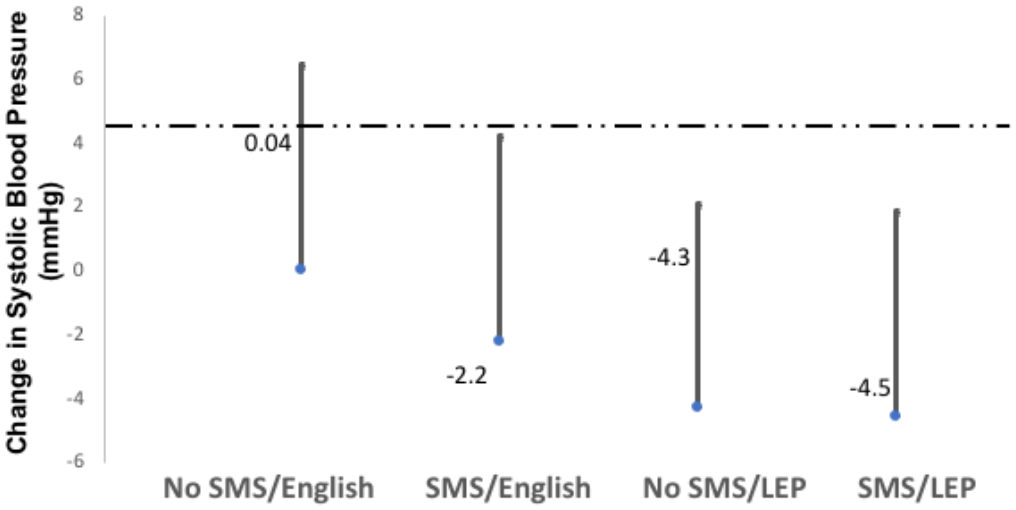
Figure 1. Mean (SD) of automated telephone self-management call duration by English proficiency. LEP: limited English proficiency.



Secondary Outcomes

Change in SBP was nonstatistically greater in the LEP group compared with the non-LEP group among patients randomized to the self-management support intervention as well as among those randomized to usual care (*P*=.74). Adjusted estimates of SBP change among patients with LEP who received the self-management support intervention were –4.5 mmHg (SD 2.49) and among patients with LEP randomized to usual care were –4.3 mmHg (SD 2.97). Participants with English proficiency randomized to the intervention had an average adjusted estimated SBP change of –2.2 mmHg (SD 3.29), whereas those randomized to usual care did not experience any change in SBP (0.04; SD 3.26; Figure 2). Similarly, patients with LEP randomized to the self-management support intervention had a nonstatistically significant greater decrease in urine albumin-to-creatinine ratio compared with their English-speaking counterparts (–72.4 mg/dL [95% CI –208.9 to 64.1] vs –11.1 mg/dL [95% –166.9 to 144.7], *P*=.29).

Figure 2. Estimated change in systolic blood pressure (95% CI) by intervention and English proficiency. LEP: limited English proficiency.



Discussion

Principal Findings

In this study of primary care patients with CKD who participated in a trial examining the impact of a comprehensive CKD self-management support program, we found that participants with LEP showed higher engagement than those with English proficiency with the HIT elements of the program by completing the automated telehealth modules more often, spending significantly longer time with each automated module, and requesting more health coach phone callbacks for further information/clarification. The LEP status did not seem to influence engagement with the more traditional health education components of the program such as language-concordant reading materials and developing at least one action plan with a health coach. These data suggest that language-concordant HIT may offer a unique opportunity to impact the health of linguistically marginalized patients. However, despite differences in engagement, the comprehensive CKD self-management support intervention had a similar null impact on change in SBP and albuminuria among both patient groups (ie, those with English proficiency and those with LEP with kidney disease).

Patient Engagement

Patient engagement has been shown to be associated with preventative behaviors such as participating in health screenings, regularly attending physician appointments, and improving dietary choices and exercise habits [12-18]. More highly engaged patients have also been shown to be less likely to smoke or consume illicit substances [25]. Further, among patients with chronic disease, higher engagement has been correlated with increased home monitoring, better treatment adherence, and more consistent medical follow-up [14,16,26-31]. Chronic diseases such as hypertension, diabetes, and kidney disease not only dominate in prevalence compared with acute diseases in the American population, but they also require consistent medical management, significant disease education, and higher levels of treatment adherence. Thus, the emphasis on patient engagement, activation, and education is paramount in our population.

Impact of Limited English Proficiency

In general, LEP is associated with poorer health status, decreased access to medical care, reduced preventive health services, and poorer quality of care [2,32-36]. Patients with LEP are more likely to have poorly controlled hypertension and worse glycemic control among those with diabetes as outpatients, as well as higher infection rates and longer hospitalizations as inpatients [37-39]. In addition, patients with LEP are more likely to have lower socioeconomic status, compounding health disparities by adding the burden of poverty, limited education, and unemployment [3]. These disparities are only partially mitigated by patient-provider language concordance or interpreter use, as LEP has been shown to be an independent predictor of poorly controlled disease [32,40,41]. With over 25 million Americans having LEP, the impact of these disparities has prompted research into innovative programs to improve health status, preventive services, and patient activation within this population [1].

Health Information Technology

Automated telephone self-management support has been highlighted by the Department of Health and Human Services as a potential tool to ameliorate health care for LEP populations, and has also been studied as an intervention in high utilizer and chronic disease cohorts [42,43]. Prior studies have shown that interventions with automated telephone self-management support can reduce hospitalizations and emergency department visits, assess medication adherence, monitor patient safety

events, and improve outcomes such as medication and appointment adherence, immunizations and screening, and patient safety triggers [44,45]. However, while some studies suggest improvements in clinical outcomes such as diabetes control, the literature is lacking on studies determining efficacy of automated telephone self-management support on management of many chronic diseases [44,46,47]. While many prior trials included patients from diverse backgrounds, subanalyses of differences between patients with English proficiency and LEP were always not conducted. To our knowledge, this is one of few studies to analyze the impact of an intervention including automated telephone self-management support between patients with LEP and those with English proficiency.

Conclusions

HIT is a burgeoning field with smartphone apps, telephone interventions, and interactive electronic medical record systems rapidly modifying the delivery of health care. The results of this study indicate that language-concordant automated telephone self-management support can be used equally well by patients with LEP compared with patients with English proficiency, with potentially even higher levels of engagement. This suggests that that language-concordant telehealth interventions could be useful in mitigating communication barriers known to negatively impact health status. This is of particular importance now, given the increase in novel telehealth modalities for care delivery in the era of COVID-19, with the number of telehealth visits more than doubling in some health systems in 2020 [48]. This surge in HIT use has also emphasized the "digital divide" where marginalized populations are less likely to have access to computers or internet services, and thus unable to participate in virtual clinic visits [49]. Because of these limitations and time needed to develop infrastructure to facilitate virtual visits, telephone encounters have become a mainstay of patient care [50].

Our findings also indicate that patients with LEP may more readily engage with language-concordant telehealth interventions as opposed to language-concordant reading materials. Reasons for this difference are likely multifactorial, but may include the fact that reading materials have to account for language, literacy, and numeracy, whereas telehealth interventions (often phone or video based) can convey similar ideas with clinical vignettes and stories, without the potential pitfalls associated with low literacy, numeracy, or even limited technological knowledge (which may contribute to lack of patient portal or web-based interventions) [23]. Thus, HIT customized to account for linguistic variation should be a key consideration in the development of novel telehealth tools.

The main limitation of this study is its small sample size, limiting power to examine the impact of the self-management support program on clinical outcome by LEP status. In addition, it was conducted in a single public health care system in California. As such, the results may not be generalizable to other settings or patient populations. However, we included a diverse patient population with a high percentage of patients with LEP.

HIT interventions are rapidly evolving to address patient needs, improve health outcomes, and increase patient engagement. As these innovations are used to increasingly engage our patients, and to the extent that the LEP population continues to grow in the United States, it is important that we develop them in a way to include, not ignore, the needs of this marginalized population. Language-concordant versions of automated telehealth and HIT tools have the potential to help bridge the disparities gap for the LEP population, and represent a unique opportunity to improve health outcomes in a disproportionately disease-burdened population.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English script for 1 week's content.
[DOCX File , 19 KB - mhealth_v9i5e24520_app1.docx]

References

1. United States Census Bureau. 2019 American Community Survey Single-Year Estimates. 2020 Sep 17. URL: https://www.census.gov/newsroom/press-kits/2020/acs-1year.html [accessed 2021-03-25]
2. DuBard CA, Gizlice Z. Language Spoken and Differences in Health Status, Access to Care, and Receipt of Preventive Services Among US Hispanics. Am J Public Health 2008 Nov;98(11):2021-2028. [doi: 10.2105/ajph.2007.119008]
3. Flores G, Abreu M, Tomany-Korman SC. Limited english proficiency, primary language at home, and disparities in children's health care: how language barriers are measured matters. Public Health Rep 2005 Aug 02;120(4):418-430 [FREE Full text] [doi: 10.1177/003335490512000409] [Medline: 16025722]
4. Wilson E, Hm Chen A, Grumbach K, Wang F, Fernandez A. Effects of limited English proficiency and physician language on health care comprehension. J Gen Intern Med 2005 Sep;20(9):800-806. [doi: 10.1111/j.1525-1497.2005.0174.x]
5. Ngo-Metzger Q, Sorkin DH, Phillips RS, Greenfield S, Massagli MP, Clarridge B, et al. Providing High-Quality Care for Limited English Proficient Patients: The Importance of Language Concordance and Interpreter Use. J GEN INTERN MED 2007 Oct 24;22(S2):324-330. [doi: 10.1007/s11606-007-0340-z]

6. Piette JD, Richardson C, Himle J, Duffy S, Torres T, Vogel M, et al. A randomized trial of telephonic counseling plus walking for depressed diabetes patients. Med Care 2011 Jul;49(7):641-648 [FREE Full text] [doi: 10.1097/MLR.0b013e318215d0c9] [Medline: 21478777]
7. AbuDagga A, Resnick HE, Alwan M. Impact of Blood Pressure Telemonitoring on Hypertension Outcomes: A Literature Review. Telemedicine and e-Health 2010 Sep;16(7):830-838. [doi: 10.1089/tmj.2010.0015]
8. Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 Apr 10;334(7600):942. [doi: 10.1136/bmj.39156.536968.55]
9. Zaheer S, Garofalo V, Rodie D, Perivolaris A, Chum J, Crawford A, et al. Computer-Aided Telephone Support for Primary Care Patients with Common Mental Health Conditions: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 10;5(4):e10224. [doi: 10.2196/10224]
10. Cox M, Basen-Engquist K, Carmack CL, Blalock J, Li Y, Murray J, et al. Comparison of Internet and Telephone Interventions for Weight Loss Among Cancer Survivors: Randomized Controlled Trial and Feasibility Study. JMIR Cancer 2017 Sep 27;3(2):e16. [doi: 10.2196/cancer.7166]
11. Piette JD, Rosland AM, Marinec NS, Striplin D, Bernstein SJ, Silveira MJ. Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically ill patients. Med Care 2013 Mar;51(3):216-223 [FREE Full text] [doi: 10.1097/MLR.0b013e318277ebf8] [Medline: 23222527]
12. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
13. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and Testing of a Short Form of the Patient Activation Measure. Health Serv Res 2005 Dec;40(6p1):1918-1930. [doi: 10.1111/j.1475-6773.2005.00438.x]
14. Greene J, Hibbard JH. Why Does Patient Activation Matter? An Examination of the Relationships Between Patient Activation and Health-Related Outcomes. J GEN INTERN MED 2011 Nov 30;27(5):520-526. [doi: 10.1007/s11606-011-1931-2]
15. Fowles JB, Terry P, Xi M, Hibbard J, Bloom CT, Harvey L. Measuring self-management of patients' and employees' health: Further validation of the Patient Activation Measure (PAM) based on its relation to employee characteristics. Patient Education and Counseling 2009 Oct;77(1):116-122. [doi: 10.1016/j.pec.2009.02.018]
16. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is Patient Activation Associated With Outcomes of Care for Adults With Chronic Conditions? Journal of Ambulatory Care Management 2007;30(1):21-29. [doi: 10.1097/00004479-200701000-00005]
17. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008 Oct(8):1-9. [Medline: 18946947]
18. Tabrizi JS, Wilson AJ, O'Rourke PK. Customer quality and type 2 diabetes from the patients' perspective: a cross-sectional study. J Res Health Sci 2010 Dec 18;10(2):69-76 [FREE Full text] [Medline: 22911927]
19. Tuot DS, Velasquez A, McCulloch CE, Banerjee T, Zhu Y, Hsu C, et al. The Kidney Awareness Registry and Education (KARE) study: protocol of a randomized controlled trial to enhance provider and patient engagement with chronic kidney disease. BMC Nephrol 2015 Oct 22;16(1):1-10. [doi: 10.1186/s12882-015-0168-4]
20. Tuot DS, Rubinsky AD, Velasquez A, McCulloch CE, Schillinger D, Handley MA, et al. Interventions to Improve Blood Pressure Control Among Socioeconomically Disadvantaged Patients With CKD: Kidney Awareness Registry and Education Pilot Randomized Controlled Trial. Kidney Med 2019;1(5):242-252 [FREE Full text] [doi: 10.1016/j.xkme.2019.07.004] [Medline: 32734204]
21. Strait A, Velasquez A, Handley MA, Leong K, Najmabadi A, Powe NR, et al. Acceptability of a multilevel intervention to improve blood pressure control among patients with chronic kidney disease in a public health care delivery system. Clin Kidney J 2018 Aug;11(4):540-548. [doi: 10.1093/ckj/sfx141] [Medline: 30094019]
22. Tuot DS, Davis E, Velasquez A, Banerjee T, Powe NR. Assessment of Printed Patient-Educational Materials for Chronic Kidney Disease. Am J Nephrol 2013;38(3):184-194. [doi: 10.1159/000354314]
23. Schillinger D, Hammer H, Wang F, Palacios J, McLean I, Tang A, et al. Seeing in 3-D: Examining the Reach of Diabetes Self-Management Support Strategies in a Public Health Care System. Health Educ Behav 2006 Aug 02;35(5):664-682. [doi: 10.1177/1090198106296772]
24. National Center for Health Statistics (Centers for Disease Control and Prevention). National Health and Nutrition Examination Survey (NHANES): Food Security Questionnaire. 2018. URL: https://wwwn.cdc.gov/nchs/data/nhanes/2017-2018/questionnaires/FCBS-CBQ-J-508.pdf [accessed 2021-04-14]
25. Salyers MP, Matthias MS, Spann CL, Lydick JM, Rollins AL, Frankel RM. The Role of Patient Activation in Psychiatric Visits. PS 2009 Nov;60(11):1535-1539. [doi: 10.1176/ps.2009.60.11.1535]
26. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do Increases in Patient Activation Result in Improved Self-Management Behaviors? Health Serv Res 2007 Aug;42(4):1443-1463. [doi: 10.1111/j.1475-6773.2006.00669.x]
27. Becker E, Roblin D. Translating primary care practice climate into patient activation: the role of patient trust in physician. Med Care 2008 Aug;46(8):795-805. [doi: 10.1097/MLR.0b013e31817919c0] [Medline: 18665059]

28. Rogvi S, Tapager I, Almdal T, Schiøtz ML, Willaing I. Patient factors and glycaemic control--associations and explanatory power. Diabet Med 2012 Oct;29(10):e382-e389. [doi: 10.1111/j.1464-5491.2012.03703.x] [Medline: 22540962]
29. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VBB, et al. Online Diabetes Self-Management Program: A randomized study. Diabetes Care 2010 Mar 18;33(6):1275-1281. [doi: 10.2337/dc09-2153]
30. Rask KJ, Ziemer DC, Kohler SA, Hawley JN, Arinde FJ, Barnes CS. Patient Activation Is Associated With Healthy Behaviors and Ease in Managing Diabetes in an Indigent Population. Diabetes Educ 2009 Apr 28;35(4):622-630. [doi: 10.1177/0145721709335004]
31. Remmers C, Hibbard J, Mosen D, Wagenfield M, Hoye R, Jones C. Is patient activation associated with future health outcomes and healthcare utilization among patients with diabetes? J Ambul Care Manage 2009;32(4):320-327. [doi: 10.1097/JAC.0b013e3181ba6e77] [Medline: 19888008]
32. Fernandez A, Schillinger D, Warton EM, Adler N, Moffet HH, Schenker Y, et al. Language Barriers, Physician-Patient Language Concordance, and Glycemic Control Among Insured Latinos with Diabetes: The Diabetes Study of Northern California (DISTANCE). J GEN INTERN MED 2010 Sep 29;26(2):170-176. [doi: 10.1007/s11606-010-1507-6]
33. Sarkar U, Karter AJ, Liu JY, Moffet HH, Adler NE, Schillinger D. Hypoglycemia is More Common Among Type 2 Diabetes Patients with Limited Health Literacy: The Diabetes Study of Northern California (DISTANCE). J GEN INTERN MED 2010 May 18;25(9):962-968. [doi: 10.1007/s11606-010-1389-7]
34. Schillinger D, Barton LR, Karter AJ, Wang F, Adler N. Does Literacy Mediate the Relationship between Education and Health Outcomes? A Study of a Low-Income Population with Diabetes. Public Health Rep 2016 Aug 02;121(3):245-254. [doi: 10.1177/003335490612100305]
35. Wisnivesky JP, Krauskopf K, Wolf MS, Wilson EA, Sofianou A, Martynenko M, et al. The association between language proficiency and outcomes of elderly patients with asthma. Annals of Allergy, Asthma & Immunology 2012 Sep;109(3):179-184. [doi: 10.1016/j.anai.2012.06.016]
36. U.S. Department of Health and Human Services Agency for Healthcare Research and Quality. National Healthcare Disparities Report. In: www.ahrq.gov/qual/qrdr10.htm. Rockville, MD: AHRQ Publication No. 11-0005; Mar 01, 2011.
37. Tang EW, Go J, Kwok A, Leung B, Lauck S, Wong ST, et al. The relationship between language proficiency and surgical length of stay following cardiac bypass surgery. European Journal of Cardiovascular Nursing 2016 Jul 09;15(6):438-446. [doi: 10.1177/1474515115596645]
38. Choi S, Lee J, Rush E. Ethnic and language disparities in diabetes care among California residents. Ethn Dis 2011;21(2):183-189. [Medline: 21749022]
39. Kim EJ, Kim T, Paasche-Orlow MK, Rose AJ, Hanchate AD. Disparities in Hypertension Associated with Limited English Proficiency. J GEN INTERN MED 2017 Feb 3;32(6):632-639. [doi: 10.1007/s11606-017-3999-9]
40. Hacker K, Choi YS, Trebino L, Hicks L, Friedman E, Blanchfield B, et al. Exploring the Impact of Language Services on Utilization and Clinical Outcomes for Diabetics. PLoS ONE 2012 Jun 4;7(6):e38507. [doi: 10.1371/journal.pone.0038507]
41. Flores G. The Impact of Medical Interpreter Services on the Quality of Health Care: A Systematic Review. Med Care Res Rev 2016 Aug 18;62(3):255-299. [doi: 10.1177/1077558705275416]
42. Quan J, Lee AK, Handley MA, Ratanawongsa N, Sarkar U, Tseng S, et al. Automated Telephone Self-Management Support for Diabetes in a Low-Income Health Plan: A Health Care Utilization and Cost Analysis. Population Health Management 2015 Dec;18(6):412-420. [doi: 10.1089/pop.2014.0154]
43. Institute of Medicine's Quality of Health Care in America project. Crossing the Quality Chasm: A New Health System for the 21st Century. In: Institute of Medicine (US) Committee on Quality of Health Care in America. Washington, DC: Washington (DC): National Academies Press (US); 2001; 2001:1-360.
44. Posadzki P, Mastellos N, Ryan R, Gunn L, Felix L, Pappas Y, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev 2016 Dec 14;12:CD009921 [FREE Full text] [doi: 10.1002/14651858.CD009921.pub2] [Medline: 27960229]
45. Lyles CR, Schillinger D, Lopez A, Handley M, Ratanawongsa N, Sarkar U. Safety Events during an Automated Telephone Self-Management Support Intervention. J Diabetes Sci Technol 2013 May;7(3):596-601. [doi: 10.1177/193229681300700303]
46. Piette JD, Weinberger M, McPhee SJ, Mah CA, Kraemer FB, Crapo LM. Do automated calls with nurse follow-up improve self-care and glycemic control among vulnerable patients with diabetes? The American Journal of Medicine 2000 Jan;108(1):20-27. [doi: 10.1016/s0002-9343(99)00298-3]
47. Lewinski AA, Patel UD, Diamantidis CJ, Oakes M, Baloch K, Crowley MJ, et al. Addressing Diabetes and Poorly Controlled Hypertension: Pragmatic mHealth Self-Management Intervention. J Med Internet Res 2019 Apr 09;21(4):e12541. [doi: 10.2196/12541]
48. Mann D, Chen J, Chunara R, Testa P, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
49. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]

50. Mehrotra A. Rapidly Converting to "Virtual Practices": Outpatient Care in the Era of Covid-19. NEJM Catalyst 2020 Apr 01:19 [FREE Full text]

Abbreviations

CKD: chronic kidney disease
IT: information technology
KARE: Kidney Awareness Registry and Education
LEP: limited English proficiency
SBP: systolic blood pressure

Edited by L Buis; submitted 22.09.20; peer-reviewed by K Wac, Z Su, AV Das; comments to author 07.12.20; revised version received 15.02.21; accepted 02.04.21; published 11.05.21.

Please cite as:
Machen L, Handley MA, Powe N, Tuot D
Engagement With a Health Information Technology–Augmented Self-Management Support Program in a Population With Limited English Proficiency: Observational Study
JMIR Mhealth Uhealth 2021;9(5):e24520
URL: https://mhealth.jmir.org/2021/5/e24520
doi:10.2196/24520
PMID:33973868

Original Paper

Influence Mechanism of the Affordances of Chronic Disease Management Apps on Continuance Intention: Questionnaire Study

Yongmei Liu[1*], PhD; Fei Jiang[1*], PhD; Peiyang Lin[1*], MSc

Business School of Central South University, Changsha City, China

[*]all authors contributed equally

Corresponding Author:
Peiyang Lin, MSc
Business School of Central South University
No 932, Lushan South Road
Yuelu District, Hunan Province
Changsha City
China
Phone: 86 15116112197
Email: linpeiyangcn@163.com

Abstract

Background: Mobile health apps are becoming increasingly popular, and they provide opportunities for effective health management. Existing chronic disease management (CDM) apps cannot meet users' practical and urgent needs, and user adhesion is poor. Few studies, however, have investigated the factors that influence the continuance intention of CDM app users.

Objective: Starting from the affordances of CDM apps, this study aimed to analyze how such apps can influence continuance intention through the role of health empowerment.

Methods: Adopting a stimulus-organism-response framework, an antecedent model was established for continuance intention from the perspective of perceived affordances, uses and gratifications theory, and health empowerment. Perceived affordances were used as the "stimulus," users' gratifications and health empowerment were used as the "organism," and continuance intention was used as the "response." Data were collected online through a well-known questionnaire survey platform in China, and 323 valid questionnaires were obtained. The theoretical model was tested using structural equation modeling.

Results: Perceived connection affordances were found to have significant positive effects on social interactivity gratification (t_{717}=6.201, P<.001) and informativeness gratification (t_{717}=5.068, P<.001). Perceived utilitarian affordances had significant positive effects on informativeness gratification (t_{717}=7.029, P<.001), technology gratification (t_{717}=8.404, P<.001), and function gratification (t_{717}=9.812, P<.001). Perceived hedonic affordances had significant positive effects on function gratification (t_{717}=5.305, P<.001) and enjoyment gratification (t_{717}=13.768, P<.001). Five gratifications (t_{717}=2.767, P=.005; t_{717}=4.632, P<.001; t_{717}=7.608, P<.001; t_{717}=2.496, P=.012; t_{717}=5.088, P<.001) had significant positive effects on health empowerment. Social interactivity gratification, informativeness gratification, and function gratification had significant positive effects on continuance intention. Technology gratification and enjoyment gratification did not have a significant effect on continuance intention. Health empowerment had a significant positive effect on continuance intention. Health empowerment and gratifications play mediating roles in the influence of affordances on continuance intention.

Conclusions: Health empowerment and gratifications of users' needs are effective ways to promote continuance intention. The gratifications of users' needs can realize health empowerment and then inspire continuance intention. Affordances are key antecedents that affect gratifications of users' needs, health empowerment, and continuance intention.

(JMIR Mhealth Uhealth 2021;9(5):e21831) doi:10.2196/21831

KEYWORDS

health empowerment; perceived affordances; uses and gratifications; S-O-R framework; continuance intention; chronic disease management app

Introduction

Background

Chronic diseases are among the main threats to people's health; they have complex causes and are difficult to cure. Self-management of health is an effective way to treat and alleviate chronic diseases. Treating chronic diseases through self-management requires individuals to have sufficient health knowledge and access to medical resources. However, the capacity of Chinese community health service institutions for chronic disease management (CDM) is weak, offline professional medical resources are difficult to obtain, and individuals lack professional guidance and know little about chronic disease treatment. With the development of internet technology and mobile phones, people are increasingly using mobile phones to manage their health. Mobile health apps are becoming increasingly popular, and they provide opportunities for effective health management [1]. However, related research has shown that existing CDM apps cannot meet users' practical and urgent needs, and user adhesion is poor [2]. A review by Triantafyllidis et al [3] found that mobile health interventions show mixed evidence in promoting the efficiency of CDM. The continuous use of CDM apps is a key means for people to manage their diseases and for enterprises to remain competitive. Most previous studies of CDM apps have focused on the design, development, or effectiveness of the apps [4,5]. Few, however, have investigated the factors that influence the continuous use intention of app users. As such, there is a need for further research on the factors that influence the continuance intention of CDM app users.

To achieve the goal of restoring health status through self-management, it is necessary to emphasize the role of health empowerment in self-management [6]. Since health empowerment emphasizes the individual's sense of control over his or her own health and the improvement of the individual's health knowledge and capabilities [7], it has a positive effect on health management [8]. Health empowerment can stimulate the individual's internal potential and health awareness; it is thus an effective strategy for helping individuals to produce healthy behaviors and maximize their level of health [9,10]. Such factors are key to continuous individual behavior. Therefore, it is feasible to explore users' continuance intention from the perspective of health empowerment. However, existing research on whether and how CDM apps can achieve health empowerment has not provided sufficient answers. Thus far, studies have found that the outcomes of health empowerment can include reinventing the self, stimulating internal strength, promoting self-management, improving quality of life, expanding social support, and promoting healthy behavior [10]. It is not yet known, however, whether health empowerment can promote continuance intention.

Meeting users' needs is an effective way to realize health empowerment and promote continuance intention [11,12]. Can CDM apps meet the need of users to achieve health empowerment and thus affect continuance intention? In recent years, researchers have increasingly adopted the uses and gratifications theory (UGT) to study continuance intention [13-15]. UGT provides theoretical guidance for us to investigate what needs information technology users have while also providing a theoretical lens for exploring how to promote users' health empowerment. Previous studies using UGT have tended to start from the gratification of users' needs while ignoring how those needs are satisfied. Affordance refers to an app's potential to support certain behaviors, emotions, and cognitions, and it can be perceived by users with relevant needs [16]. When users feel that their needs are supported by technology products, their own needs are met. Therefore, this study used the theoretical perspective of affordances (technical features) to explain how a CDM app can meet users' different needs. It is also feasible to study the empowerment of technology products by considering their technical features [17]. Further, the gratification of user needs can explain the psychological mechanism through which technical products empower users.

In summary, starting from the affordances of CDM apps, this study aimed to analyze how such apps can influence users' continuance intention through the role of health empowerment. This study addresses the following 2 questions: First, how do CDM apps meet users' needs? Second, how do users' need gratifications and health empowerment affect users' continuance intention?

Theoretical Foundation

The stimulus-organism-response (S-O-R) framework was used as the basic framework for the model in this study. Mehrabian and Russell [18] noted that an individual in a specific environment will be stimulated by various factors that will trigger internal changes in the individual, which will lead to a corresponding response by the individual. In the context of this study, we combined affordance, UGT, and health empowerment, using affordances as stimulus variables, user need gratifications and health-empowerment awareness as organism variables, and continuance intention as the response variable.

Affordances are the technical features of CDM apps (environmental stimulus), which provide environmental factors for us to examine how CDM apps affect user need gratification, health empowerment, and continuance intention. Perceived affordance, as an environmental stimulus attribute, is a novel perspective for studying individual behavior [19,20]. At the same time, affordance also provides research ideas for examining how user needs are met. We summarized the comprehensive affordances of CDM apps through 3 kinds of affordances (ie, utilitarian, connective, hedonic) according to Scheepers and Middleton [21].

Menon [22] argued that health empowerment is the individual's perception of control over his or her health and health care; it reflects the internalization of health ideals and goals on personal and social levels. Londoño and Schulz [8] applied psychological empowerment to the field of health care to obtain a measure of mental health empowerment comprising 4 dimensions: self-determination, meaning, impact, and competence.

UGT mainly focuses on identifying users' psychological needs when selecting media, and it explains users' subsequent psychological and behavioral changes according to the degree of gratification [23]. Previous research has used UGT to explore

users' continuance intentions mainly in terms of 4 types of gratification: utilitarian, social, enjoyment, and technology gratification [12]. An important factor in CDM app use is whether the app has suitable functions to support users' health management (function gratification). Therefore, this study added the unique gratification of "self-management function gratification" as an aspect of utilitarian gratification. The explanation of all constructs of UGT is shown in Table 1.

Table 1. The constructs of uses and gratifications theory (UGT).

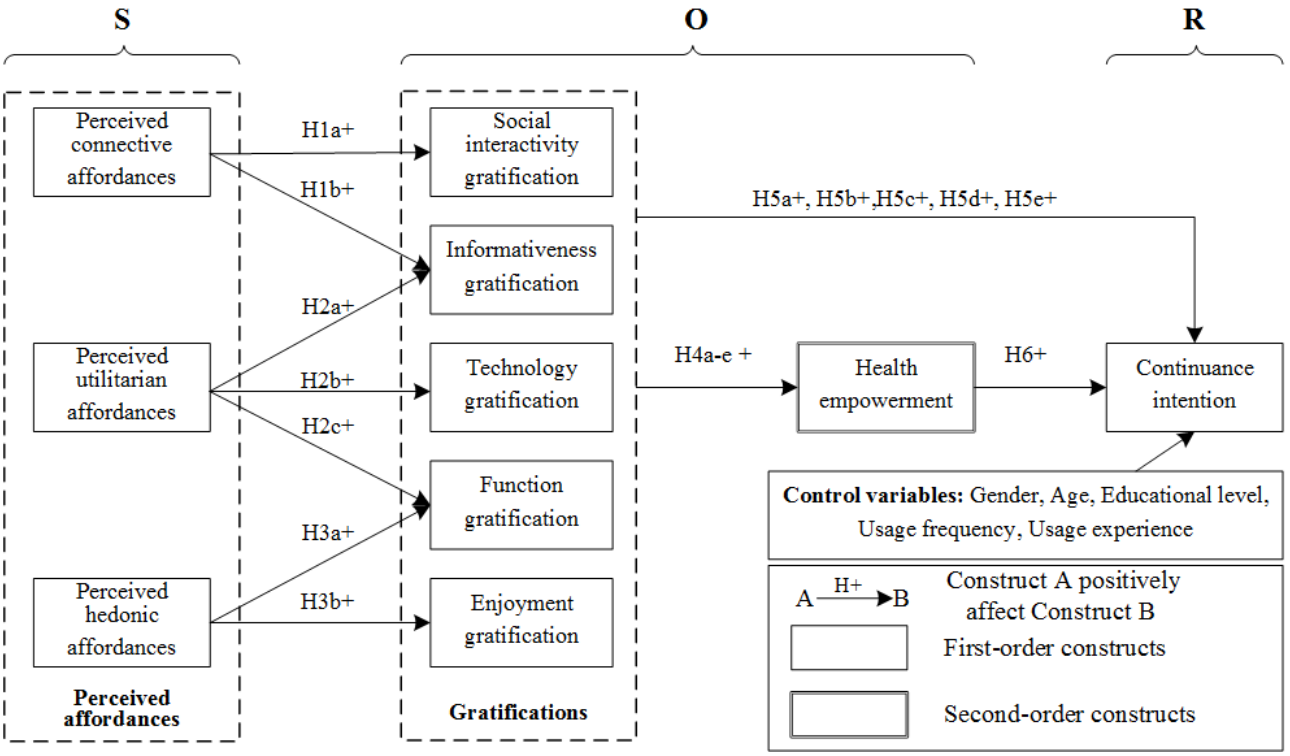
Gratifications	Construct	Definition in this research
Social	Social interactivity gratification	The degree to which the CDM[a] app assists a user to establish and maintain contact with other users. For example, through the app, a user can get to know other users (doctors or peers) and interact with emotions and information.
Utilitarian	Informativeness gratification	The degree to which using the CDM app helps manage information or deliver meaning. For example, through the various multimedia modes of the app, one can obtain health knowledge or exchange authentic experiences.
Utilitarian	Function gratification	The extent to which users perceive the CDM app to help them manage diseases and solve problems. For example, CDM apps can guide user management, help develop personalized plans, and acquire knowledge.
Enjoyment	Enjoyment gratification	The degree to which users believe that the CDM app brings relaxation and relief. For example, users can access health resources and other users through the app, thereby improving self-efficacy and alleviating depression.
Technology	Technology gratification	The degree to which the CDM apps enables users to perform health management in a convenient and easy way. For example, the friendly interface and ubiquity of the app allow users to conveniently manage chronic diseases.

[a]CDM: chronic disease management.

Research Model and Hypotheses

Figure 1 shows the research model. When users perceive the different affordances of CDM apps, they will have different degrees of gratification according to how well their needs are supported. Based on different degrees of gratification, users will further make judgments about their health empowerment and willingness to continue using the CDM app. The hypotheses of this study are H1-H6, and detailed information about the hypotheses is shown in Multimedia Appendix 1.

Figure 1. Research model.



Methods

Measurements

This study's theoretical model has 10 key constructs, and all measurements refer to previously verified scales. The research model contains 3 types of perceived affordances and 5 UGT-related constructs. They are the user's evaluation of the information technology characteristics of CDM apps and the perception of gratifications after use. Their measurements come from scales that have been verified in the field of information systems. The measurement of 3 kinds of perceived affordances referred to research by Tang and Zhang [24] on social e-commerce. Perceived connective affordances refer to how a user perceives the CDM app allows them to interact with other users, and the measurement contains 3 items (eg, "The CDM app has some features that support me to communicate with others"). Perceived utilitarian affordances refers to the action possibility that a user perceives the CDM app helps them improve health management, and the measurement contains 4 items (eg, "The CDM app has some features that support me to collect information related to health management"). Perceived hedonic affordances refers to a user's perception that the CDM app can make them feel relaxed and relieve anxiety, and the measurement contains 3 items (eg, "The CDM app has some features that that support my relaxation spirit").

Combining users' practical needs with UGT, we measured 5 user-need gratifications: social interactivity, informativeness, technology, function, and enjoyment. For the measurement instruments of social interactivity, informativeness, and enjoyment gratification, we referred to research by Kim et al [14] on mobile social network sites [14]. Social interactivity gratification contains a total of 4 items (eg, "The CDM app is satisfying me in helping me maintain social relationships with other users."). Informativeness gratification contains 4 items (eg, "I am satisfied with using the CDM app to search for various health information."). Enjoyment gratification contains 4 items (eg, "I am in a good mood when using the CDM app to conduct health management–related activities."). We followed the process by Gan and Li [12] to measure technology gratification. The measurement by Gan and Li appears to be related to a study about a type of social media app and contains 3 items (eg, "A CDM app is the simplest and most cost-effective way to manage health."). Moreover, we followed the process by Glasgow et al [25] to measure function gratification, and their measurement appears to be related to a study about chronic illness care. The scale includes 3 items (eg, "The CDM app is satisfying me in helping me customize the treatment plans.").

We adopted the measurement instruments of Londoño and Schulz [8] to measure health empowerment. The measurement by Londoño and Schulz appears to be related to a study on asthma and showed good reliability and validity. With health empowerment as a second-order reflexive construct in the research model, each first-order factor has 3 items: meaning (eg, "Using the CDM app to meet my health needs is of unique significance to me."), competence (eg, "I have mastered some of the necessary skills to meet my health needs by using the CDM app."), self-determination (eg, "I have great autonomy in how to manage health in the process of using the CDM app,"), impact (eg, "My impact on how to manage my own health is great in the process of using the CDM app.").

Finally, a scale with 3 items was used to measure continuance intention, which referred to the measurement instruments used by Hooi and Cho [26]. Multimedia Appendix 2 lists all the scales used in this study.

Since this study used a Chinese CDM app, the original English questionnaire was translated into Chinese by professional researchers with consideration of the specifics of the app to ensure equivalence in the translation [12]. After the preliminary preparation of the questionnaire, several researchers were invited to review it. Based on their feedback, items were reworded to improve readability and clarity. A 5-point Likert scale was used for all measurements ("disagree very much" to "agree very much"). We also measured relevant demographic information and variables that might affect the dependent variables, including gender, age, education, income, usage experience, and use frequency [27].

Data Collection

Data were collected online through a well-known questionnaire survey platform in China [28]. This is a professional online survey platform with more than 40 million members and more than 2.6 million sample members. It covers all areas of society and ensures the authenticity of the sample; it has been used and confirmed by a large number of scholars in different fields [23,29]. At the same time, the platform has a sample database of CDM app users, and it issues questionnaires for such users. To further identify suitable responders, we set up some screening questions at the beginning of the questionnaire (eg, "Have you used a related CDM app [Micro Sugar, Hypertension Housekeeper, Diabetes Control, etc.]" "Have you communicated with other users of the CDM app?" "Have you obtained health knowledge through the CDM app?"). Only participants who pass all the screening questions can continue to fill in the questionnaire; otherwise, the ability to answer questions will be terminated. Data collection lasted 1 week, during which a total of 382 samples were collected. Among them, 27 cases were unfinished, accounting for 7.1% of the total (27/382). To ensure the reliability of the questionnaire, we eliminated 19 responses (19/382, 5.0%) for which the questionnaires took less than 5 minutes to complete and 13 responses (13/382, 3.4%) for which the answers consistently showed the same or similar values (which appeared as unlikely responses). Finally, 323 valid questionnaires were obtained.

Cohen power tables were used to test whether the sample size was sufficient to detect the effects of interest [30]. This method is recommended for the power analysis of studies using partial least squares (PLS) [31]. In this study, the largest number of independent variables of the construct (health empowerment and continuance intention) was 5. According to Cohen power tables, assuming a medium e ect size (f^2=0.15), statistical power of 0.80, and significance level of .05, the smallest sample size was 91. Therefore, a sample size of 323 was sufficient in our study.

Data Analysis

A PLS structural equation model was used for data analysis, employing SmartPLS 3.0. PLS is particularly suitable for dealing with multistage models [32]. The measurement model was evaluated through confirmatory factor analysis (CFA); then, the structural model was tested by bootstrapping [33].

Results

General Statistical Description

Among the 323 valid questionnaires, the proportions of men and women were 45.0% (145/323) and 55.0% (178/323), respectively. Among the participants, the 31-40-year age group was the largest, accounting for 43.3% (140/323); 68.1% (267/323) of the participants had 1-3 years of experience using a chronic disease app; and 88.9% (287/323) of the participants used it less than 3 times a day. Regarding the level of education, 78.3% (253/323) of the participants had a bachelor's degree. Table 2 shows the detailed demographic information.

Table 2. Demographic characteristics (n=323).

Characteristics	Results, n (%)
Gender	
Male	145 (45.0)
Female	178 (55.0)
Age (years)	
18-30	136 (42.1)
31-40	140 (43.3)
41-50	30 (9.3)
51-60	17 (5.3)
App use experience (years)	
0-1	47 (14.6)
1-2	118 (36.5)
2-3	102 (31.6)
3-4	39 (12.1)
4-5	13 (4.0)
>5	4 (1.2)
App use frequency (times/day)	
0-3	287 (88.9)
4-6	28 (8.7)
≥7	8 (2.5)
Educational level	
High school or below	9 (2.8)
Junior college	37 (11.5)
Undergraduate	253 (78.3)
Masters or above	24 (7.4)
Monthly income (RMB)	
<3000	28 (8.7)
3000-4999	44 (13.6)
5000-7999	92 (28.5)
8000-9999	74 (22.9)
10,000-14,999	69 (21.4)
≥15,000	16 (5.0)

Measurement Model

First, we tested the reliability and validity of the constructs in the model through CFA. The factor loadings of all first-order constructs were greater than the recommended value of 0.7 [34]. Health empowerment was a second-order reflexive construct in which the dimension of self-determination was less than the recommended value of 0.7, thus excluding the dimension of self-determination [35]. The minimum values for composite reliability and Cronbach α were 0.827 and 0.715, respectively, which are greater than the recommended value of 0.7, thus ensuring the reliability of the construct measurement [36]. In addition, the average variance extracted (AVE) was higher than the critical value of 0.5 [36]. In summary, these indicators confirm the convergent validity of the measurement model. Multimedia Appendix 3 provides the test results of convergence validity and reliability.

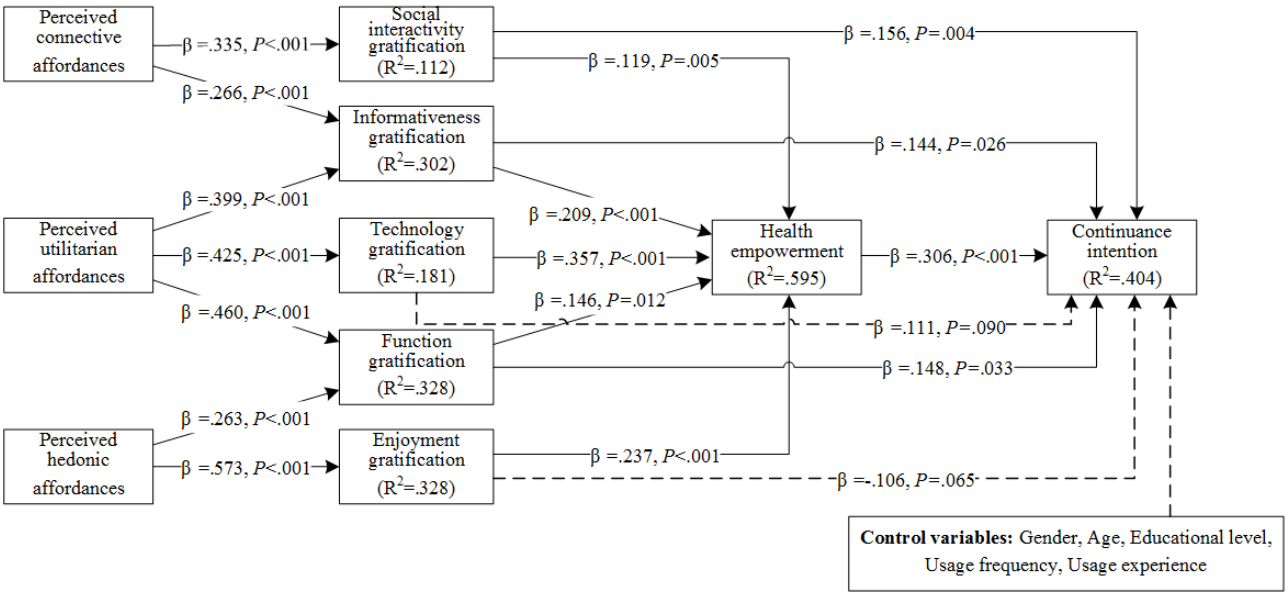
Second, we tested discriminant validity by comparing the square root of the AVE of the constructs with the correlation coefficient between constructs. If the former is greater than the latter, discriminant validity can be proven [37]. The correlations between the constructs were all less than 0.7, and the values on the diagonal (the square root of AVE) were significantly greater than the correlation coefficients between constructs. Therefore, discriminant validity was verified. For details, see Multimedia Appendix 4.

Finally, according to the correlation matrix between constructs, the correlation coefficient between technology gratification and health empowerment was the maximum value (0.608), which is less than the recommended value of 0.8, and the maximum value of all variance inflation factors was 1.793, well below the critical value of 10 [38]. Thus, multicollinearity was not a serious problem in this study. The Harman single-factor test was performed using SPSS version 22 (IBM Corp, Armonk, NY) to analyze method bias. The analysis was performed without rotation, and no factor was specified. If the number of factors extracted is more than one and the variance contribution rate of the first factor does not exceed 40%, it is generally considered that common-method bias is not serious [39]. In this study, the obtained first-factor variance contribution rate was 24.23% (less than 40%); thus, common-method bias was not a serious problem.

Structural Model

Next, we tested the hypotheses. Figure 2 shows the results. Perceived connection affordances were found to have significant positive effects on social interactivity gratification and informativeness gratification. Thus, H1a and H1b are supported. Perceived utilitarian affordances had significant positive effects on informativeness gratification, technology gratification, and function gratification; thus, H2a, H2b, and H2c are supported. Perceived hedonic affordances had significant positive effects on function gratification and enjoyment gratification. Therefore, H3a and H3b are supported. The gratification of the user's 5 needs had a significant positive effect on health empowerment. Thus, H4a, H4b, H4c, H4d, and H4e are supported. Social interactivity, informativeness, and function gratification had significant positive effects on continuance intention. Thus, H5a, H5b, and H5d are all supported. Technology gratification and enjoyment gratification did not have a significant effect on continuance intention. Therefore, H5c and H5e are not supported. Health empowerment had a significant positive effect on users' continuance intention. H6 is therefore supported. Finally, all control variables did not have significant effects on the dependent variable. Overall, the model explained 59.5% of the variation in health empowerment and 40.4% of the variation in continuance intention and thus had good explanatory power.

Figure 2. Analysis results of the structural model.



Mediating Effect Analysis of Health Empowerment

The bootstrap method was used for the analysis of mediating effects. Compared to traditional methods, the advantage of this method is that it can directly test the indirect effects of independent variables on the dependent variables, and it does not require the mediating effects to follow a normal distribution [40,41]. The sample size was selected as 5000, and the selected confidence interval method was the bias-corrected and

accelerated bootstrap. Table 3 summarizes the mediating analysis results obtained using SmartPLS 3.2.8. The effect of perceived connection affordances on continuance intention worked through 4 paths. Therefore, there is evidence that health empowerment played a mediating role in the influence of perceived connection affordances, social interactivity gratification, and informativeness gratification on continuance intention. Perceived utilitarian affordances affected continuance intention through 5 paths. Thus, there is evidence that health empowerment played a mediating role in the influence of perceived utilitarian affordances, informativeness gratification, technology gratification, and function gratification on continuance intention. Perceived hedonic affordances affected continuance intention through 3 paths. Thus, there is evidence that health empowerment mediated the influence of perceived hedonic affordances, function gratification, and enjoyment gratification on continuance intention.

Table 3. Bootstrapping results.

Indirect effect path	95% confidence interval	Path of the corresponding direct effect	95% confidence interval	Type of mediation
PCA[a] - SIG[b] - CI[c]	0.016 to 0.097	PCA - CI	0.069 to 0.175	Partial
PCA -SIG - HE[d] - CI	0.003 to 0.027	PCA - CI	0.069 to 0.175	Partial
PCA - IG[e] - CI	0.006 to 0.075	PCA - CI	0.069 to 0.175	Partial
PCA - IG - HE - CI	0.007 to 0.034	PCA - CI	0.069 to 0.175	Partial
PUA[f] - IG - CI	0.005 to 0.118	PCA - CI	0.190 to 0.338	Partial
PUA - IG – HE - CI	0.011 to 0.048	PCA - CII	0.190 to 0.338	Partial
PUA - TG[g] - HE -CI	0.022 to 0.081	PCA - CI	0.190 to 0.338	Partial
PUA - FG[h] - CI	0.006 to 0.138	PCA - CI	0.190 to 0.338	Partial
PUA - FG - HE - CI	0.004 to 0.049	PCA - CI	0.190 to 0.338	Partial
PHA[i] - FG - CI	0.005 to 0.088	PCA - CI	–0.041 to 0.110	full
PHA - FG - HE - CI	0.003 to 0.030	PCA - CI	–0.041 to 0.110	full
PHA - EG[j] - HE - CI	0.018 to 0.074	PCA - CI	–0.041 to 0.110	full

[a]PCA: perceived connection affordances.
[b]SIG: social interactivity gratification.
[c]CI: continuance intention.
[d]HE: health empowerment.
[e]IG: informativeness gratification.
[f]PUA: perceived utilitarian affordances.
[g]TG: technology gratification.
[h]FG: function gratification.
[i]PHA: perceived hedonic affordances.
[j]EG: enjoyment gratification.

Discussion

Principal Findings

This study obtained the following important findings: First, we found that CDM apps can achieve health empowerment by gratifying users' needs through affordances. Gratification is key to health empowerment. Among the affordances of CDM apps, perceived connective, utilitarian, and hedonic affordances help gratify needs related to the connection (social interactivity, informativeness), utilitarian experience (informativeness, self-management, technology), and hedonic experience (enjoyment, function), respectively. Health empowerment gives users a sense of a self-determined environment for generating healthy behaviors. Therefore, it is important for the user's health that the user understand how the CDM app promotes his or her health empowerment. This finding is the same as those in the studies by Nelson et al [17] and Audrain-Pontevia and Menvielle [42], once again proving that the application of information technology in the health field can effectively promote health empowerment.

Second, health empowerment was found to be a key factor influencing users' continuance intention, and it mediated the influences of affordance and gratification on continuance intention. This result suggests that enhancing the user's sense of empowerment is crucial to the lifespan of CDM apps. Promoting the user's sense of health empowerment is a feasible way to stimulate continuance intention. Different from previous research on health empowerment, which focused on the rights of patients [8-10], this finding shows that health empowerment can also promote the continued use of health-related applications. Promoting health empowerment is a win-win for patients and app providers.

To suit our research scenario, we added "the gratification of self-management functions." We found that gratification from the self-management function was a major factor influencing users' health empowerment and continuance intention. Therefore, CDM apps should provide health management functions corresponding to users' needs to help users manage their own health, thereby improving user satisfaction with the functional design. We also found that technology gratification and enjoyment gratification did not have a significant effect on continuance intention; thus, H5c and H5e were not supported. The possible reason for H5c being rejected is that technology gratification is more concerned with convenience and ease of use; such factors affect the initial adoption of the app but have little effect on continuance intention. Regarding H5e, this study was conducted in the context of CDM; thus, the demand for enjoyment might not be so large relative to other user needs. In addition, when we used bootstrapping to test the mediating effect, we found that its effect on continuance intention was fully mediated by health empowerment.

Implications for Research

This study's theoretical contributions are as follows: First, by combining affordance, UGT, and health empowerment, an antecedent model for the intention to continue using CDM apps was established. This provides novel research ideas and a research framework for investigating information system use continuance, supplementing existing research on this topic. This study considered the role of health empowerment as a contextual factor. Unlike previous studies that usually explore the effect of perceived usefulness and satisfaction on continued use intention [43,44], this study focused on CDM apps' ability to empower users and health empowerment's effect on continuance intention. This provides a new perspective for exploring the continued use of health applications.

Second, this study provides empirical guidance for studying how to promote health empowerment. It also enriches the literature on factors influencing health empowerment. Considering the important role of health empowerment in stimulating individual health behaviors, this study found that satisfying the different needs of users has a positive effect on users' health empowerment. Although past research has indicated that health empowerment can be achieved by satisfying individual needs [11,45], this has never been tested empirically in the context of CDM app usage. Future research may consider starting from the perspective of "affordance-gratification-health empowerment" to explore more empowerment methods.

Third, this study expands the application of UGT in information systems literature by considering the role of a new type of gratification and also deepens our understanding of the antecedents of continuance intention. UGT has usually been used to explain why people choose a particular media without involving a CDM app [12,46]. This study found that the gratification of self-management functions was an important factor influencing the continuous use of CDM apps. Further prior research using UGT was based on the gratification of various user needs and ignored the antecedents of gratifications [13-15]. This study revealed how a CDM app can promote the gratification of user needs from the perspective of affordance. Therefore, this study expands UGT research and deepens our understanding of it.

Limitations and Future Research

This study has some limitations. First, the samples involved Chinese CDM app users. The conclusions should be carefully considered in other scenarios. Comparative cross-cultural and cross-platform studies could be considered in the future. Second, this study considered 3 kinds of affordances for CDM apps based on previous studies. There may be other types of affordances that could be considered; future research could develop more affordances. In addition to the 4 kinds of gratifications scholars have used in the past, this study uniquely added "gratification of self-management needs." Users may have other types of needs, which should be considered in future research. Third, the data in this study are cross-sectional data, and there is a lack of analysis of influencing factors based on the time span. Longitudinal research can be considered in future studies. Finally, although we have specified the target of the questionnaire and set up screening items, some false respondents may be included in the final sample. Future studies should consider research methods that have more control over the respondents (eg, experimental designs).

Conclusion

Under the overall framework of S-O-R, this study examined the influence mechanism of CDM app users' continuance intention based on the perspective of perceived affordance, UGT, and health empowerment. The results indicated that users' perceptions of an app's affordances can promote the gratification of needs, and the gratification of key needs (ie, social interactivity, informativeness, technology, and function gratification) can stimulate users' continuance intention. At the same time, the gratification of users' needs can promote users' cognitions of health empowerment, thus stimulating continuance intention. Health empowerment was found to play a mediating role in the influence of gratification on continuance intention. From a practical perspective, app service providers should design apps from the perspective of social interaction (eg, providing social networks), utilitarian functions (eg, health self-management), and hedonic functions (eg, enhancing the user's interest). By meeting users' various needs, app developers can improve the user's ability to control his or her own health, thus achieving the purpose of extending the life of the app.

Acknowledgments

This work was funded in part by the National Natural Science Foundation of China (Grant Numbers 71771219, 72071213). In addition, we appreciate the academic committee in the "Mobile Health" Ministry of Education- China Mobile Joint Laboratory reviewed the project proposal and gave us the ethical approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The detailed information of the hypotheses.
[DOCX File , 43 KB - mhealth_v9i5e21831_app1.docx]

Multimedia Appendix 2
Measurement instruments.
[DOCX File , 28 KB - mhealth_v9i5e21831_app2.docx]

Multimedia Appendix 3
The results of convergence validity and reliability.
[DOCX File , 43 KB - mhealth_v9i5e21831_app3.docx]

Multimedia Appendix 4
Discriminant validity.
[DOCX File , 30 KB - mhealth_v9i5e21831_app4.docx]

References

1. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
2. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile Applications for Control and Self Management of Diabetes: A Systematic Review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
3. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]
4. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
5. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
6. Ramsay Wan C, Vo L, Barnes CS. Conceptualizations of patient empowerment among individuals seeking treatment for diabetes mellitus in an urban, public-sector clinic. Patient Educ Couns 2012 Jun;87(3):402-404. [doi: 10.1016/j.pec.2011.09.010] [Medline: 22005710]
7. Cerezo PG, Juvé-Udina ME, Delgado-Hito P. Concepts and measures of patient empowerment: a comprehensive review. Rev Esc Enferm USP 2016;50(4):667-674 [FREE Full text] [doi: 10.1590/S0080-623420160000500018] [Medline: 27680054]
8. Londoño AMM, Schulz PJ. Influences of health literacy, judgment skills, and empowerment on asthma self-management practices. Patient Educ Couns 2015 Jul;98(7):908-917. [doi: 10.1016/j.pec.2015.03.003] [Medline: 25907737]
9. Aujoulat I, Marcolongo R, Bonadiman L, Deccache A. Reconsidering patient empowerment in chronic illness: a critique of models of self-efficacy and bodily control. Soc Sci Med 2008 Mar;66(5):1228-1239. [doi: 10.1016/j.socscimed.2007.11.034] [Medline: 18155338]
10. Aujoulat I, d'Hoore W, Deccache A. Patient empowerment in theory and practice: polysemy or cacophony? Patient Educ Couns 2007 Apr;66(1):13-20. [doi: 10.1016/j.pec.2006.09.008] [Medline: 17084059]
11. Anderson RM, Funnell MM, Butler PM, Arnold MS, Fitzgerald JT, Feste CC. Patient empowerment. Results of a randomized controlled trial. Diabetes Care 1995 Jul;18(7):943-949. [doi: 10.2337/diacare.18.7.943] [Medline: 7555554]
12. Gan C, Li H. Understanding the effects of gratifications on the continuance intention to use WeChat in China: A perspective on uses and gratifications. Computers in Human Behavior 2018 Jan;78:306-315. [doi: 10.1016/j.chb.2017.10.003]
13. Lee HE, Cho J. What Motivates Users to Continue Using Diet and Fitness Apps? Application of the Uses and Gratifications Approach. Health Commun 2017 Dec;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
14. Kim MJ, Lee C, Contractor NS. Seniors' usage of mobile social network sites: Applying theories of innovation diffusion and uses and gratifications. Computers in Human Behavior 2019 Jan;90:60-73. [doi: 10.1016/j.chb.2018.08.046]
15. Hsu MH, Chang CM, Lin HC, Lin YW. Determinants of continued use of social media: the perspectives of uses and gratifications theory and perceived interactivity. Information Research 2015;20(2):671 [FREE Full text]

16. Norman DA. The Psychology of Everyday Things. New York, NY: Basic Books; 1988.
17. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: An empirical smart wristband study. Computers in Human Behavior 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]
18. Mehrabian A, Russell JA. An approach to environmental psychology. Cambridge, MA: The MIT Press; 1974.
19. Seet B, Goh T. Exploring the affordance and acceptance of an e-reader device as a collaborative learning system. The Electronic Library 2012 Aug 03;30(4):516-542. [doi: 10.1108/02640471211252229]
20. Zhao Y, Liu J, Tang J, Zhu Q. Conceptualizing perceived affordances in social media interaction design. AP 2013 Mar;65(3):289-303. [doi: 10.1108/00012531311330656]
21. Scheepers R, Middleton C. Personal ICT Ensembles and Ubiquitous Information Systems Environments: Key Issues and Research Implications. CAIS 2013;33:381-392. [doi: 10.17705/1cais.03322]
22. Menon ST. Toward a model of psychological health empowerment: implications for health care in multicultural communities. Nurse Educ Today 2002 Jan;22(1):28-39; discussion 40. [doi: 10.1054/nedt.2001.0721] [Medline: 11886225]
23. Huang J, Zhou L. Timing of web personalization in mobile shopping: A perspective from Uses and Gratifications Theory. Computers in Human Behavior 2018 Nov;88:103-113. [doi: 10.1016/j.chb.2018.06.035]
24. Tang J, Zhang P. The impact of atmospheric cues on consumers' approach and avoidance behavioral intentions in social commerce websites. Computers in Human Behavior 2020 Jul;108:105729. [doi: 10.1016/j.chb.2018.09.038]
25. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). Med Care 2005 May;43(5):436-444. [doi: 10.1097/01.mlr.0000160375.47920.8c] [Medline: 15838407]
26. Hooi R, Cho H. Virtual world continuance intention. Telematics and Informatics 2017 Dec;34(8):1454-1464. [doi: 10.1016/j.tele.2017.06.009]
27. Coursaris CK, Yun Y, Sung J. Twitter Users vs. Quitters: A Uses and Gratifications and Diffusion of Innovations Approach in Understanding the Role of Mobility in Microblogging. 2010 Presented at: Ninth International Conference on Mobile Business and Ninth Global Mobility Roundtable (ICMB-GMR); June 13-15, 2010; Athens, Greece. [doi: 10.1109/icmb-gmr.2010.44]
28. wjx.cn. Changsha Ranxing Information Technology Co., Ltd. URL: https://www.wjx.cn/ [accessed 2021-05-01]
29. Lien CH, Cao Y. Examining WeChat users' motivations, trust, attitudes, and positive word-of-mouth: Evidence from China. Computers in Human Behavior 2014 Dec;41:104-111. [doi: 10.1016/j.chb.2014.08.013]
30. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
31. Benitez J, Henseler J, Castillo A, Schuberth F. How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. Information & Management 2020 Mar;57(2):103168. [doi: 10.1016/j.im.2019.05.003]
32. Fornell C, Bookstein FL. Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. Journal of Marketing Research 2018 Dec 20;19(4):440-452. [doi: 10.1177/002224378201900406]
33. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
34. Carmines EG, Zeller RA. Reliability and validity assessment. Thousand Oaks, CA: Sage Publications, Inc; 1979.
35. Shin JI, Chung KH, Oh JS, Lee CW. The effect of site quality on repurchase intention in Internet shopping through mediating variables: The case of university students in South Korea. International Journal of Information Management 2013 Jun;33(3):453-463. [doi: 10.1016/j.ijinfomgt.2013.02.003]
36. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
37. Fornell C, Larcker DF. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. Journal of Marketing Research 1981 Aug;18(3):382-388. [doi: 10.1177/002224378101800313]
38. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis, 7th edition. Hoboken, NJ: Pearson Prentice Hall; 2010.
39. Podsakoff PM, Organ DW. Self-Reports in Organizational Research: Problems and Prospects. Journal of Management 1986;12(4):531-544. [doi: 10.1177/014920638601200408]
40. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/brm.40.3.879] [Medline: 18697684]
41. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002 Mar;7(1):83-104 [FREE Full text] [doi: 10.1037/1082-989x.7.1.83] [Medline: 11928892]
42. Audrain-Pontevia A, Menvielle L. Do online health communities enhance patient-physician relationship? An assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018 Aug;31(3):154-162. [doi: 10.1177/0951484817748462] [Medline: 29280679]
43. Beldad AD, Hegner SM. Expanding the Technology Acceptance Model with the Inclusion of Trust, Social Influence, and Health Valuation to Determine the Predictors of German Users' Willingness to Continue using a Fitness App: A Structural Equation Modeling Approach. International Journal of Human–Computer Interaction 2017 Nov 30;34(9):882-893. [doi: 10.1080/10447318.2017.1403220]

44. Wu B. Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
45. Anderson RM, Funnell MM, Fitzgerald JT, Marrero DG. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. Diabetes Care 2000 Jun;23(6):739-743 [FREE Full text] [doi: 10.2337/diacare.23.6.739] [Medline: 10840988]
46. Heravi A, Mubarak S, Raymond Choo K. Information privacy in online social networks: Uses and gratification perspective. Computers in Human Behavior 2018 Jul;84:441-459. [doi: 10.1016/j.chb.2018.03.016]

Abbreviations

AVE: average variance extracted
CDM: chronic disease management
CFA: confirmatory factor analysis
PLS: partial least squares
S-O-R: stimulus-organism-response
UGT: uses and gratifications theory

Edited by L Buis; submitted 26.06.20; peer-reviewed by R Cruz Martínez, H Kondylakis; comments to author 15.09.20; revised version received 09.11.20; accepted 08.04.21; published 13.05.21.

Please cite as:
Liu Y, Jiang F, Lin P
Influence Mechanism of the Affordances of Chronic Disease Management Apps on Continuance Intention: Questionnaire Study
JMIR Mhealth Uhealth 2021;9(5):e21831
URL: https://mhealth.jmir.org/2021/5/e21831
doi:10.2196/21831
PMID:33983126

Original Paper

Efficacy of Zemedy, a Mobile Digital Therapeutic for the Self-management of Irritable Bowel Syndrome: Crossover Randomized Controlled Trial

Melissa Hunt[1], PhD; Sofia Miguez[1], BA; Benji Dukas[1], BA; Obinna Onwude[2], MD; Sarah White[3], PhD

[1]Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

[2]Bold Health Limited, UK, London, United Kingdom

[3]Population Health Research Institute, St George's University of London, London, United Kingdom

Corresponding Author:
Melissa Hunt, PhD
Department of Psychology
University of Pennsylvania
425 S University Ave
Levin Building
Philadelphia, PA, 19104-6018
United States
Phone: 1 2158986478
Email: mhunt@psych.upenn.edu

Abstract

Background: Patients with irritable bowel syndrome (IBS) experience abdominal pain, altered bowel habits, and defecation-related anxiety, which can result in reduced productivity and impaired health-related quality of life (HRQL). Cognitive behavioral therapy (CBT) has been shown to reduce symptoms of IBS and to improve HRQL, but access to qualified therapists is limited. Smartphone-based digital therapeutic interventions have potential to increase access to guided CBT at scale, but require careful study to assess their benefits and risks.

Objective: The aim of this study was to test the efficacy of a novel app, Zemedy, as a mobile digital therapeutic that delivers a comprehensive CBT program to individuals with IBS.

Methods: This was a crossover randomized controlled trial. Participants were recruited online and randomly allocated to either immediate treatment (n=62) or waitlist control (n=59) groups. The Zemedy app consists of 8 modules focusing on psychoeducation, relaxation training, exercise, the cognitive model of stress management, applying CBT to IBS symptoms, reducing avoidance through exposure therapy, behavioral experiments, and information about diet. Users interact with a chatbot that presents the information and encourages specific plans, homework, and exercises. The treatment was fully automated, with no therapist involvement or communication. At baseline and after 8 weeks, participants were asked to complete the battery of primary (Irritable Bowel Syndrome Quality of Life [IBS-QOL], Gastrointestinal Symptom Rating Scale [GSRS]) and secondary (Fear of Food Questionnaire [FFQ], Visceral Sensitivity Index [VSI], Gastrointestinal Cognition Questionnaire [GI-COG], Depression Anxiety Stress Scale [DASS], and Patient Health Questionnaire-9 [PHQ-9]) outcome measures. Waitlist controls were then offered the opportunity to crossover to treatment. All participants were assessed once more at 3 months posttreatment.

Results: Both intention-to-treat and completer analyses at posttreatment revealed significant improvement for the immediate treatment group compared to the waitlist control group on both primary and secondary outcome measures. Gains were generally maintained at 3 months posttreatment. Scores on the GSRS, IBS-QoL, GI-COG, VSI, and FFQ all improved significantly more in the treatment group ($F_{1,79}$=20.49, P<.001, Cohen d=1.01; $F_{1,79}$=20.12, P<.001, d=1.25; $F_{1,79}$=34.71, P<.001, d=1.47; $F_{1,79}$=18.7, P<.001, d=1.07; and $F_{1,79}$=12.13, P=.001, d=0.62, respectively). Depression improved significantly as measured by the PHQ-9 ($F_{1,79}$=10.5, P=.002, d=1.07), and the DASS Depression ($F_{1,79}$=6.03, P=.02, d=.83) and Stress ($F_{1,79}$=4.47, P=.04, d=0.65) subscales in the completer analysis but not in the intention-to-treat analysis. The impact of treatment on HRQL was mediated by reductions in catastrophizing and visceral sensitivity.

Conclusions: Despite its relatively benign physical profile, IBS can be an extraordinarily debilitating condition. Zemedy is an effective modality to deliver CBT for individuals with IBS, and could increase accessibility of this evidence-based treatment.

Trial Registration: ClinicalTrials.gov NCT04170686; https://www.clinicaltrials.gov/ct2/show/NCT04170686

(JMIR Mhealth Uhealth 2021;9(5):e26152) doi:10.2196/26152

KEYWORDS

digital health; irritable bowel syndrome; cognitive behavioral therapy; CBT; efficacy; mHealth; self-management; IBS; randomized controlled trial; app

Introduction

Background

Irritable bowel syndrome (IBS) is a chronic gastrointestinal (GI) disorder of multifactorial etiology that is characterized by abnormal centralized pain processing. IBS is defined by recurrent abdominal pain occurring at least one day per week in the last 3 months, associated with two or more of the following: related to defecation, associated with changes in the frequency or form of bowel movements (ie, characterized by constipation, diarrhea, or an alternating mix of the two). IBS is highly prevalent, affecting up to 10% of the US population. Many studies have demonstrated that IBS has high rates of psychiatric comorbidity (up to 90% in treatment-seeking patients) [1,2], and causes social and occupational impairment [3]. Beyond the core symptoms of abdominal pain and altered bowel habits, individuals with IBS suffer from a host of related difficulties that substantially impair health-related quality of life (HRQL) and functioning. Visceral hypersensitivity, common among IBS patients, is a phenomenon in which people feel normal gut sensations that most people would be unaware of, and experience many of these sensations as more painful compared with healthy controls [4]. Anxiety and visceral hypersensitivity are highly correlated [5]. Anxiety and hypervigilance related to the sensations exacerbate the hypersensitivity [6].

Illness-related anxiety is high among patients with IBS, and is a better predictor of impairment in quality of life than actual symptom severity [7]. A major component of this anxiety is "catastrophizing," in which individuals envision the worst possible outcome of their GI symptoms and in turn develop maladaptive coping strategies [3]. Catastrophizing is highly correlated with impairment in HQRL for patients with IBS [8]. Because of their catastrophizing, many individuals with IBS engage in significant avoidance behavior that can easily meet the diagnostic criteria for agoraphobia [9].

Cognitive Behavioral Therapy for IBS

Over the past two decades, cognitive behavioral therapy (CBT) has repeatedly proven to be an efficacious treatment for individuals suffering from IBS [10,11]. Specifically, there is empirical support that CBT reduces GI symptom severity and impairment in quality of life [12,13]. These CBT treatments typically include components of psychoeducation about the brain-gut axis, mindfulness and relaxation training [14], reducing automatic negative thoughts related to GI catastrophizing [15], exposure therapy to feared and avoided sensations and situations [16], and reducing visceral hypersensitivity [12]. One meta-analysis including 20 psychological treatments for IBS found that GI cognition change and GI-specific anxiety were important mediators in improving GI-related quality of life and GI symptom severity [17].

Although CBT is a promising treatment, access to IBS-specific CBT remains low for patients. There is a lack of clinicians competent in delivering GI-specific CBT [3]. Additionally, the cost of treatment looms high; individuals often lack insurance coverage for psychotherapy and must pay out of pocket, which can be burdensome given the hundreds of dollars their IBS likely already costs them [18]. It is therefore necessary to develop a cheaper, more easily accessible alternative mode of treatment.

Many groups have tested variants of CBT for IBS with limited or distant therapist involvement (eg, via email) [15,19] and typically obtain robust effect sizes. These studies generally showed that web-based and telephone-based CBT improved IBS more than treatment as usual (eg, [20]). Several treatment manuals and self-help books are available that detail the CBT treatment approach, and one [21] was found to be efficacious as a stand-alone treatment in a randomized controlled trial (RCT) [22].

In today's digitized world, the mobile health (mHealth) industry is growing. The industry is currently valued at close to US $50 billion and is expected to multiply by nearly five times over the next decade [23]. Thousands of mobile apps exist to improve health across the spectrum. Mobile apps have multiple advantages, including low cost, privacy, accessibility, and convenience for the user.

CBT is among the forms of treatment increasingly being delivered via apps. In their review of eight CBT apps, Rathbone et al [24] found that CBT self-help apps can be efficacious, most notably in alleviating depressive symptoms. They also cited the willingness of participants to engage in therapy as a key component of the apps' success [24]. A component of many mHealth apps, and specifically those that use CBT, is automated guidance and feedback. Automated guidance has been found to be effective in reducing substance abuse among urban women and emerging adults [25,26]. Kelders et al [27] compared an automated treatment for depression with standard, in-person clinical treatment and found that depressive symptoms were moderately reduced for those in the automated group, although not as strongly as found for the in-person treatment group. However, Mason and Andrews' [28] internet CBT study found that "specialist assessments and initial face-to-face contact do not influence treatment outcome, and that patients do just as well with an automated assessment." Hauser-Ulrich et al [29] developed a smartphone app to treat chronic pain through CBT. This app employs a chatbot that guides users through modules [29]. In their RCT, the authors found improvements in pain-related impairment, pain intensity, and general well-being

for those who used the app for 8 weeks [29]. Thus, there is strong evidence to suggest that automated treatment in a CBT app may be highly effective in delivering integrative behavioral health care for patients with disorders at the boundary between physical symptoms and psychological distress.

Aim

As self-help modalities are increasingly available online and through smartphone apps, it is important to test the efficacy of those apps through rigorous, controlled research. The purpose of this study was to test the efficacy of a novel digital app (Zemedy) that applies CBT to IBS.

Methods

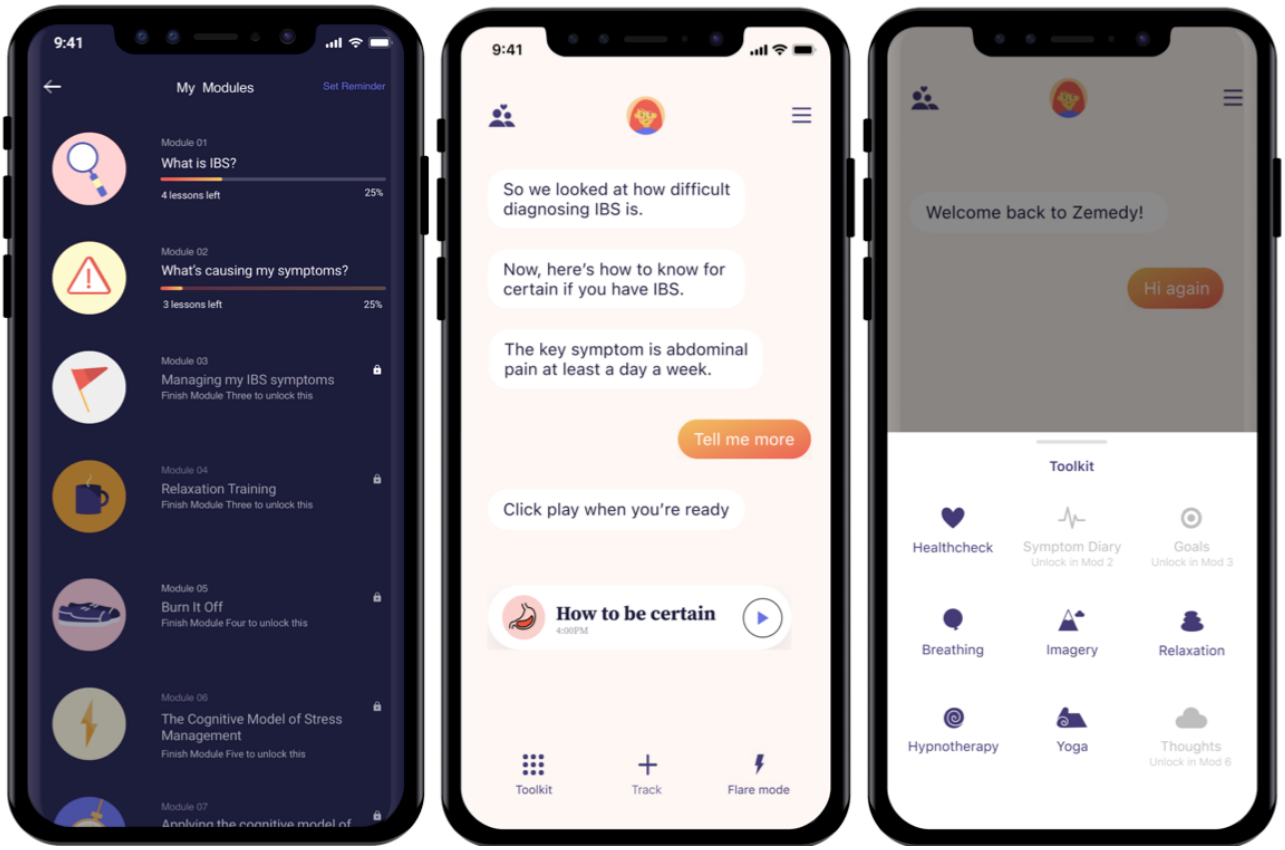
Trial Registration

This study was approved by the Institutional Review Board of the University of Pennsylvania. All participants provided electronic consent prior to participation in the study. The deidentified dataset analyzed in the study is available from the corresponding author upon request. This trial was registered at ClinicalTrials.gov as NCT04170686.

App Description

Zemedy 1.0 is a mobile phone app designed by Bold Health, a UK-based digital health company, in collaboration with the principal investigator (MH) based on her empirically supported self-help book [21]. The app treats IBS through CBT specifically developed for the condition. Users of either iOS or Android smartphones are guided through the app by a chatbot with whom they "text." The app consists of 10 modules. The first two modules are devoted to psychoeducation about the etiology of IBS and CBT's effectiveness in treating it. The remaining eight modules teach users about various CBT strategies to mitigate the impact of IBS on daily life, including relaxation training, exercise, cognitive restructuring and decatastrophizing, exposure exercises to reduce avoidance, and behavioral experiments. It also encourages a healthy (but not highly restrictive) diet. See Figure 1 for screenshots of the Zemedy app. Users are prompted to apply these strategies to their daily lives. Similar to the chronic pain treatment developed by Hauser-Ulrich et al [29], Zemedy is designed to be completed in 8 weeks. Participants were encouraged to read through the first 5 modules (education, relaxation training, and exercise) in the first week, and to practice relaxation exercises daily. The remaining modules were designed to be worked through approximately one per week, with practice and homework exercises performed daily to learn and apply the skills.

Figure 1. Screenshots of Zemedy.



The app also includes a "flare module," which users can access at any point to address immediate GI pain and anxiety. Shah et al [14] found that mind-body interventions such as relaxation training and hypnosis have moderate effect sizes in reducing IBS symptoms. The flare module contains a variety of exercises such as deep breathing, progressive muscle relaxation, relaxation imagery, and hypnotherapy scripts that help mitigate distress and discomfort in the moment.

Participants were provided with a link to download the app. They were provided the app at no cost. The entire intervention was delivered within the app with no human involvement (eg, therapist guidance or feedback). If participants experienced technical difficulties, they could reach out to technical support at Bold Health. They received a single email at 4 weeks from a research coordinator in the trial providing general encouragement to continue working through the app (if they were in the immediate treatment group) or to "hang in there" (if they were in the waitlist control group).

Design

This was a randomized waitlist control crossover trial with assessments performed at baseline, postintervention (8 weeks), following crossover to intervention for the waitlist control group, and at follow-up (3 months postintervention). After completing the consent and all baseline measures, participants were randomly allocated by a research coordinator to either the immediate treatment or waitlist control group using the coin toss function of random.org. After 8 weeks, all participants were asked to complete the same battery of measures. At that point, participants in the waitlist group were crossed over and were given access to the app. After 8 weeks of access, they were asked to complete the battery of questionnaires again. All participants were then assessed one final time 3 months after completing the treatment phase. Unfortunately, the onset of the COVID-19 pandemic coincided with the follow-up portion of the trial, and it is unclear the extent to which the pandemic affected both attrition and long-term results.

Sample Size

The power analysis showed that 30 participants per randomized group would have 85% statistical power at a two-sided significance level of .05 to detect an effect size of 0.76. The effect size was chosen as previous studies of internet-delivered CBT for IBS reported similar effect sizes for HRQL and GI symptoms outcomes [15,22]. Assuming 50% attrition, as is common in internet-based intervention studies, we aimed to recruit 120 participants to have ample power to detect main effects, and to explore potential mediators and moderators.

Inclusion and Exclusion Criteria

Inclusion criteria consisted of being 18 years of age or older, and reporting having been previously diagnosed by a physician with IBS or meeting the Rome IV criteria by self-report. We did not specify a time frame for the physician diagnosis; therefore, it is possible that some participants were originally diagnosed under Rome III criteria. Owning a smartphone and computer/internet literacy were de facto eligibility criteria.

Exclusion criteria consisted of reporting having received a diagnosis of another comorbid GI disorder such as celiac disease or an inflammatory bowel disease. Exclusion criteria also included severe depression and/or suicidal ideation, defined as a score of 20 or above on the Patient Health Questionnaire-9 (PHQ-9), and/or positive endorsement of active suicidal ideation or intent on a separate suicide question. Twenty-five individuals met this criterion. They were excluded from the trial, but were given immediate access to the app. The principal investigator, who is a licensed clinical psychologist, followed up with each of these individuals to complete a risk assessment and offer other resources such as referrals to local in-person providers.

Participants

A total of 146 potential participants were screened, 121 of whom met the inclusion criteria. Participants were recruited for the trial through IBS-specific social media sites with a combination of graphic advertisements, and posts and comments on threads informing site users about the study. Most participants came to the study through Facebook (n=30), Twitter (n=32), and the IBS subReddit (n=51). There were no face-to-face components to the trial in terms of recruitment, assessment, or intervention. Posts and advertisements included a link to a secure University of Pennsylvania Qualtrics study page. On following the link, potential participants would first see the detailed explanation of the research (consent form; see Multimedia Appendix 1) and would consent to completing the baseline questionnaires. Questionnaires were completed via Qualtrics and could be downloaded securely by the research team. Participants were identified by email during data collection. All data were stored in a deidentified format. All recruitment and follow-up occurred between October 1, 2019 and November 1, 2020. The trial ended upon successful completion.

All but five participants reported that they had been diagnosed with IBS by a physician at some point (which could have been under Rome III or Rome IV criteria). The five participants who did not report a physician diagnosis all met stringent Rome IV criteria by self-report. Of the 30 (24.8%) participants who reported a physician diagnosis but did not meet stringent Rome IV criteria, 7 reported pain 3 days a month and would have met Rome III criteria. Another 5 participants reported even less frequent pain. Four women reported pain solely during their menstrual period. Four people failed to meet the duration criteria (less than 6 months total since onset). The final 10 participants met only one, rather than two, of the three criteria beyond frequent abdominal pain.

With respect to IBS subtype, 48 (39.7%) participants reported diarrhea-predominant IBS, 28 (23.1%) reported mixed subtype, 11 (9.1%) reported constipation-predominant IBS, and 4 (3.3%) reported undifferentiated IBS. The remaining 30 (who did not meet all Rome IV criteria) were not divided into subtypes.

Randomization and Blinding

Participants who met the inclusion criteria were allocated to condition using the coin toss feature of random.org. A total of 62 participants were assigned to the immediate treatment condition and 59 were assigned to the waitlist control. The allocation sequence was concealed to participants until they were enrolled, had completed baseline data collection, and had been assigned to a group. The majority of baseline symptom severity measures were not significantly different between the immediate treatment and waitlist control groups. However, participants in the waitlist control group reported significantly more depression and more impaired HRQL than those in the immediate treatment group. Although the design should have yielded a low risk of bias from randomization, the slight differences in symptom severity at baseline suggest some

concerns about randomization, according to the Cochrane risk of bias tool [30].

Because of the nature of the trial (immediate treatment versus waitlist control group), neither participants nor research coordinators were blinded to condition. However, there were no deviations from the intended intervention. Moreover, all outcome data were self-reported. Thus, blinding of evaluators was neither possible nor necessary.

Procedure

Participants in the immediate treatment group were given the link to access the Zemedy app, and were encouraged to download it and begin working through the modules immediately. The waitlist control group was told they would be given access to the app in 8 weeks. Four weeks after baseline, participants in the treatment group were emailed to encourage them to continue using the app, and the waitlist control group was emailed to offer encouragement, remind them they were still enrolled in the study, and let them know that they would be receiving the follow-up questionnaires in 4 weeks.

Eight weeks after completing the baseline questionnaire, all participants were emailed a link to a second Qualtrics page that contained all of the same measures as completed at baseline. Those in the waitlist control condition were then given access to the app.

After having had access to the app for 8 weeks, participants in the waitlist control group were emailed a link to the third battery of questionnaires that was identical to the battery received by the treatment group after 8 weeks of app usage, which included the same measures as contained in the baseline battery.

Finally, all participants were emailed a final link to the last battery of questionnaires (again identical to the battery at baseline and posttreatment) 3 months after they completed the active treatment phase. Upon completion of each round of questionnaires, participants received US $20 in Amazon credit.

If at any point a participant had indicated a significant increase in depressive symptoms or the onset of suicidal ideation, the team would have alerted the principal investigator (a licensed clinical psychologist) who would have reached out to that individual to perform a risk assessment and offer referrals to local resources. No such adverse events occurred.

Primary Outcome Measures

IBS-Quality of Life

The IBS-Quality of Life (IBS-QOL) questionnaire is a 34-item self-report measure specific to IBS designed to assess the impact of IBS on quality of life [31,32]. The IBS-QOL has high internal consistency (Cronbach α=.95), high reproducibility (intraclass correlation coefficient=0.86), and good construct validity [32]. Qualitative score ranges are 0-31 (minimal or mild), 32-66 (moderate), and 67-100 (severe impairment).

Gastrointestinal Symptom Rating Scale-IBS

The Gastrointestinal Symptom Rating Scale-IBS (GSRS-IBS) contains 13 self-report items rated on a 6-point Likert scale ranging from 1 (no discomfort at all) to 7 (very severe discomfort) [33]. Total scores range from 0 to 78. The GSRS-IBS has 5 subscales, including abdominal pain, bloating, constipation, diarrhea, and satiety. Each dimension has demonstrated high internal consistency, with Cronbach α ranging from .74 (pain) to .85 (satiety). Furthermore, the GSRS-IBS has demonstrated both high test-retest reliability, with intraclass correlations among the factors ranging from 0.55 (pain) to 0.70 (bloating), as well as high construct validity [33]. Overall internal consistency was good in our sample with Cronbach α=.81. The GSRS has been used as a primary outcome measure in several recent RCTs of IBS treatments [9], and the Rome Foundation reports that it is shorter and more user-friendly than the IBS Severity Scoring System [34]. Qualitative score ranges are 0-20 (minimal or mild), 21-39 (moderate), and 40-78 (severe).

Secondary Outcome Measures

Rome IV Criteria Questionnaire

We used a questionnaire to determine whether participants met the current Rome IV diagnostic criteria for IBS. Our questionnaire was based on the Rome IV IBS-Specific Questionnaire, which is a validated self-report scale that covers the diagnostic criteria for IBS. It has been found to have acceptable sensitivity and high specificity as well as good test-retest reliability [35]. We used a modified, shortened version with 10 items that covered all diagnostic criteria.

Fear of Food Questionnaire

The Fear of Food Questionnaire (FFQ) is an 18-item self-report questionnaire that measures fear, avoidance of food, as well as life interference and loss of pleasure from eating [36]. Items are rated on a Likert scale ranging from 0 (not at all) to 5 (absolutely). It has excellent internal consistency reliability with Cronbach α=.96 and strong 2-week test-retest reliability at r=0.93, $P<.001$ [36]. The FFQ also shows good criterion and known-groups validity [36]. Qualitative score ranges are 0-15 (minimal), 16-30 (mild), 31-45 (moderate), and 46-90 (severe).

Visceral Sensitivity Index

The Visceral Sensitivity Index (VSI) is a unidimensional 15-item scale that measures GI symptom–specific anxiety [6,37]. Items are rated on a Likert scale ranging from 0 (strongly disagree) to 5 (strongly agree). It has high internal consistency (Cronbach α=.93) and a mean interitem correlation of 0.47 [37,38]. It has good criterion, construct, and predictive validity [6]. Qualitative score ranges are 0-10 (minimal or mild), 11-30 (moderate), and 31-75 (severe).

Gastrointestinal Cognitions Questionnaire

The Gastrointestinal Cognitions Questionnaire (GI-COG) consists of 16 self-report items that are rated on a 5-point Likert scale, ranging from 0 (hardly) to 4 (very much). Individual items are summed and total scores range from 0 to 64. The questionnaire consists of three subscales: the pain/life interference subscale (eg, “When I feel my GI symptoms acting up, I’m afraid the pain will be excruciating and intolerable”), the social anxiety subscale (eg, “If I have to get up and leave an event, meeting, or social gathering to go to the bathroom people will think there’s something wrong with me”), and the

disgust sensitivity subscale (eg, "The thought of fecal incontinence is terrifying. If it happened, it would be awful"). The GI-COG has been shown to have excellent internal consistency (Cronbach α=.92) and test-retest reliability (r=0.87, *P*=.001) [39]. Qualitative score ranges are 0-19 (minimal or mild), 20-39 (moderate), and 40-64 (severe).

Depression Anxiety Stress Scale

The Depression Anxiety Stress Scale (DASS) is a 42-item self-administered questionnaire that measures the magnitude of depression, anxiety, and stress independently. Internal consistency for each of the subscales of the questionnaire are high, with Cronbach α of .96 to .97 for DASS-Depression, .84 to .92 for DASS-Anxiety, and .90 to .95 for DASS-Stress [40,41]. The DASS has been found to be a highly reliable and valid measure of the constructs it is intended to assess [42].

PHQ-9 Assessment

The PHQ-9 is a depression scale that consists of 9 self-report items. The 9 items aim to quantify the 9 criteria upon which the diagnosis of depressive disorders is based in the Diagnostic and Statistical Manual of Mental Disorders-IV. The PHQ-9 can establish a depressive disorder diagnosis and depressive symptom severity [43]. Each of the 9 items can be scored from 0 (not at all) to 3 (nearly every day); therefore, scores can range from 0 to 27. The PHQ-9 has been found to demonstrate high internal reliability, with Cronbach α of .89 when tested in a primary care setting and .86 when tested in an obstetrics-gynecology setting [43].

Dose

Dosage was measured according to the number of modules completed. The mobile app sent usage data to the backend system each time a participant visited the app. Data include the time and date of each session on the app.

Statistical Analysis

Univariate general linear models in SPSS V25 were used to examine between-group effects at posttreatment (8 weeks), controlling for baseline levels of the dependent variable. Paired-sample *t* tests were used to examine within-group changes over the treatment phase for each group and maintenance of gains from posttreatment to 3-month follow-up. The robustness of these analyses was examined in an intention-to-treat sensitivity analysis using multiple imputation. As shown below, missing data at follow-up were not entirely missing at random. Therefore, baseline outcome measures were included in the imputation model as predictors together with the follow-up set of measures with missing data and imputation using the fully conditional specification [44] performed to create 15 imputed datasets. Regression models were then fitted as in the primary analysis, and pooled estimates of the treatment effect were calculated. Three sets of imputed datasets were created, one for each follow-up data point, with baseline measures included in each.

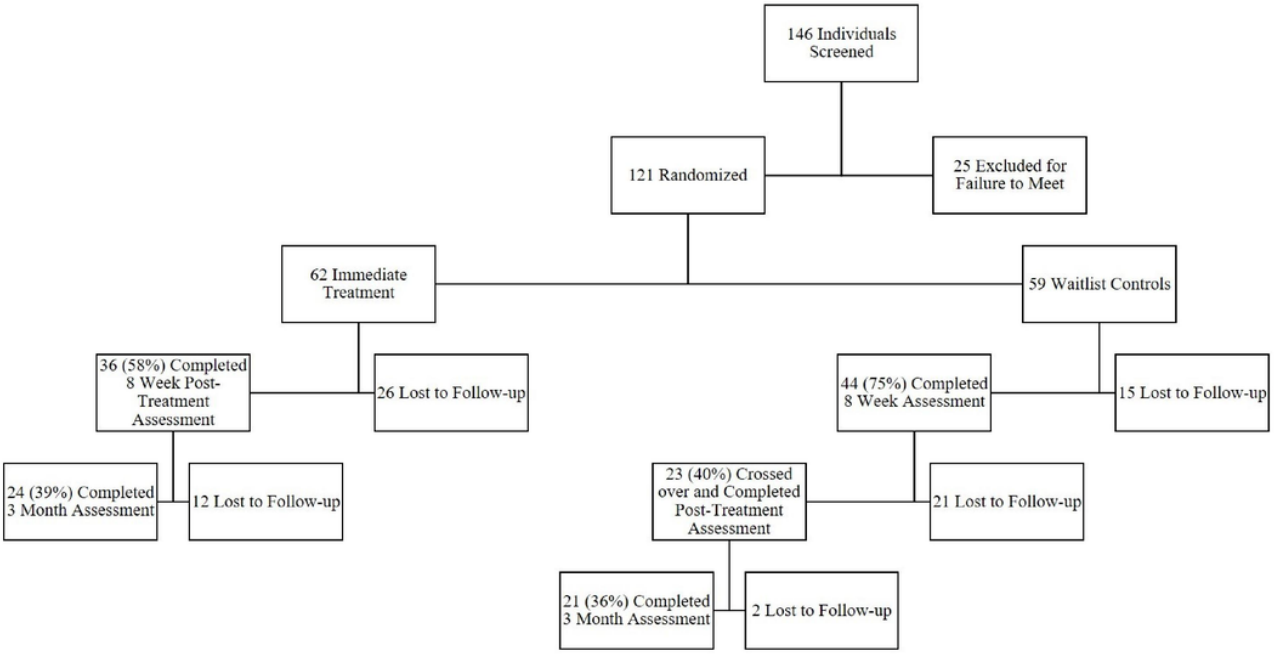
Change in visceral anxiety, catastrophizing, and fear of food (calculated as the change from baseline to 8 weeks) were explored as possible mediators of GI symptoms and quality of life at 8 weeks using regression analysis with estimates of indirect effects calculated using a percentile bootstrap estimation approach with 5000 samples implemented with the PROCESS macro Version 3.5 [45]. Both direct and indirect effects are reported. The direct effect quantifies the estimated difference in the dependent variable (GI symptoms or quality of life) between two cases that are equal on the mediator but differ by one unit on treatment assignment (ie, intervention vs waitlist group). The indirect effect quantifies how much two cases, one assigned to immediate treatment and the other to waitlist, are estimated to differ on the dependent variables (GI symptoms or quality of life) as a result of the treatment's influence on the mediator, which in turn influences the dependent variable. Two sets of models were fitted: the first tested the mediator variables separately with simple mediator models, and the second fitted a parallel mediator model where the three mediators were tested simultaneously. The baseline level of the dependent variable was included as a covariate in all mediation models.

Results

Participant Characteristics

The mean participant age was 32 years (SD 10.2, range 18-63). Of the total 121 participants, 76.0% (n=92) were white, 5.8% (n=7) were Hispanic, 5.0% (n=6) were black, 4.1% (n=5) were Asian, and the remaining 9.1% (n=11) identified as mixed race or other. With respect to gender, 75.2% (91/121) identified as female and 24.8% (30/121) identified as male. With respect to marital status, 43.0% (52/121) reported being single, 32.2% (39/121) reported being married, 19.0% (23/121) reported having a partner or cohabiting, and 5.8% (7/121) reported being divorced at baseline. With respect to employment, 22.3% (27/121) of the participants were students, 15.7% (19/121) reported being employed part time, 47.9% (58/121) reported being employed full time, and 14.0% (17/121) reported that they were not working when completing the baseline surveys. See Figure 2 for the CONSORT diagram of participant flow through the study.

Figure 2. CONSORT diagram of participant flow through the study.



There were no significant differences between the immediate treatment and waitlist control groups on any of the demographic variables, or in baseline GI symptoms, visceral sensitivity, catastrophizing, or fear of food. However, as noted above, the waitlist control group was found to be slightly more distressed than the treatment group at baseline. The waitlist control group reported significantly more depression (PHQ-9, t_{119}=2.99, *P*=.003; DASS-Depression t_{119}=2.11, *P*=.04) and more impaired quality of life (t_{119}=2.04, all *P*=.04) than the immediate treatment group, although effect sizes were modest (*d*=0.38 for the DASS, *d*=0.54 for PHQ-9, and *d*=0.37 for IBS-QOL). Thus, baseline symptoms were controlled in all analyses.

There were no univariate outliers found at baseline.

Outcomes

Completer analyses assessing the impact of treatment on outcome at 8 weeks revealed significant improvement for the immediate treatment group, relative to the waitlist control group, for both primary outcomes of GI symptom severity and HRQL ($F_{1,79}$=20.12, *P*<.001, Cohen *d*=1.02 and $F_{1,79}$=20.49, *P*<.001, *d*=1.25, respectively). With respect to the secondary outcome measures, GI-specific catastrophizing, visceral anxiety, and fear of food all improved significantly more in the treatment group ($F_{1,79}$=34.71, *P*<.001, *d*=1.47; $F_{1,79}$=18.7, *P*<.001, *d*=1.07; and $F_{1,79}$=12.13, *P*=.001, *d*=0.62, respectively). Finally, depression improved significantly more in the immediate treatment group as measured by both the PHQ-9 ($F_{1,79}$=10.5, *P*=.002, *d*=1.07), and the DASS Depression ($F_{1,79}$=6.03, *P*=.02, *d*=0.83) and Stress ($F_{1,79}$=4.47, *P*=.04, *d*=0.65) subscales. Only the DASS Anxiety subscale failed to show a significant advantage for the treatment group ($F_{1,79}$=1.84, *P*=.18, *d*=0.41). See Table 1 for all means and SDs across all assessment timepoints. These results were replicated in the intention-to-treat analyses using multiple imputation, with the exception of the PHQ-9 and DASS scores, which were nonsignificant (Table 2).

For the immediate treatment group, all of the outcome variables changed significantly from pretreatment to posttreatment with the exception of the DASS Depression subscale, which showed only marginally significant improvement (Table 3). Sensitivity analysis using multiple imputation found the same pattern of significance.

Table 1. Mean (SD) outcome measures across the trial for the immediate treatment and waitlist control groups.

Outcome	Baseline		Eight weeks		Waitlist post-treatment (n=23)	Three months	
	Immediate treatment (n=62)	Waitlist control (n=59)	Immediate treatment (n=36)	Waitlist control (n=44)		Immediate treatment (n=24)	Waitlist control (n=21)
IBS-QOL[a]	53.63 (18.67)	60.48 (18.29)	34.25 (19.78)	58.19 (18.53)	76.6 (20.07)	38.08 (18.42)	43.98 (21.1)
GSRS[b]	36.76 (12.77)	37.75 (12.02)	27.56 (10.12)	38.18 (10.79)	34.26 (14.98)	27.83 (9.37)	30.95 (11.88)
GI-COG[c]	36.92 (13.35)	40.07 (12.04)	22.44 (13.72)	40.84 (11.23)	33.3 (12.34)	23.75 (12.06)	31.71 (14.11)
VSI[d]	51.74 (12.29)	53.54 (11.44)	38.14 (16.21)	53.57 (12.37)	46.43 (12.78)	41.08 (14.13)	45.00 (12.63)
FFQ[e]	52.87 (19.14)	55.46 (18.21)	41.22 (22.23)	53.75 (18.08)	46.10 (19.87)	42.83 (20.99)	42.38 (19.87)
PHQ[f]	8.32 (5.29)	11.03 (4.66)	5.78 (4.20)	10.32 (4.29)	10.30 (5.80)	6.92 (5.71)	10.33 (5.97)
DASS[g]							
Depression	11.65 (9.88)	15.59 (10.69)	7.83 (7.88)	15.45 (10.39)	14.43 (10.89)	9.08 (8.26)	16.38 (12.89)
Stress	17.84 (9.56)	18.71 (8.97)	12.72 (8.65)	18.82 (9.99)	18.78 (10.03)	15.08 (8.40)	16.86 (9.69)
Anxiety	12.03 (7.35)	12.19 (9.14)	8.67 (6.38)	12.05 (9.72)	11.83 (9.72)	9.08 (7.76)	10.00 (6.99)

[a]IBS-QOL: IBS Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Table 2. Significance of treatment allocation at 8 weeks using multiple imputation.

Measure	*t* value	*P* value (two-tailed)
IBS-QOL[a]	–2.8	.005
GSRS[b]	–2.8	.005
GI-COG[c]	–3.4	.001
VSI[d]	–2.8	.006
FFQ[e]	–2.4	.02
PHQ[f]	–1.7	.08
DASS[g]		
Depression	–1.3	.21
Stress	–1.2	.25
Anxiety	–0.6	.55

[a]IBS-QOL: IBS Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Table 3. Improvement from baseline to posttreatment for the immediate treatment group.

Measure	*t* value (*df*=35)	*P* value (two-tailed)
IBS-QOL[a]	4.368	<.001
GSRS[b]	3.312	.002
GI-COG[c]	5.603	<.001
VSI[d]	3.454	.001
FFQ[e]	3.523	.001
PHQ[f]	2.327	.03
DASS[g]		
Depression	1.707	.10
Stress	2.273	.03
Anxiety	2.164	.04

[a]IBS-QOL: IBS Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Clinically Significant Change

In terms of *clinically* significant change, we used Criterion B (falling within 2 SD of the healthy mean), which is more conservative than Criterion A (falling 2 SD below the pathological mean) [46]. For GI symptoms, the mean GSRS score for healthy controls is 12 (SD 11), leading to a cut-off point of 34. In the immediate treatment group, 24 out of 36 participants (66%) met this criterion at posttreatment. For HRQL, the mean IBS-QOL score for healthy controls is 5 (SD 11), leading to a cut-off point of 27. In the immediate treatment group, 16 out of 36 participants (44%) met this criterion at posttreatment. From another perspective, the qualitative range for minimal to mild impairment on the IBS-QOL is 0-31. An additional 2 participants would meet this slightly less stringent criterion, leading to a total of 50% of participants in the immediate treatment group showing an excellent response. This yields a number needed to treat of 2.

After completing the 8-week follow-up questionnaires, the waitlist group was crossed over to active treatment and was given access to the app for 8 weeks. Paired-samples *t* tests comparing their scores at the initial 8-week follow-up to their scores posttreatment revealed significant improvement in HRQL, catastrophizing, visceral anxiety, and fear of food, but not on GI symptoms, depression, or anxiety (Table 4). Sensitivity analysis using multiple imputation showed a similar pattern of significance at the 5% level but with lower *P* values closer to the significance level.

Table 4. Improvement in the waitlist control group after crossover to active treatment.

Measure	*t* value (*df*=22)	*P* value (two-tailed)
IBS-QOL[a]	–3.124	.005
GSRS[b]	–1.308	.20
GI-COG[c]	–2.748	.01
VSI[d]	–2.618	.02
FFQ[e]	–3.509	.002
PHQ[f]	0.103	.92
DASS[g]		
Depression	–1.537	.14
Stress	–0.361	.72
Anxiety	0.360	.72

[a]IBS-QOL: IBS Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Three-month follow-up data were collected for all participants (both the immediate treatment group and the waitlist group who had been crossed over to treatment) between March and October of 2020. Unfortunately, this meant that all follow-up data were collected after the onset of the COVID-19 pandemic. Nevertheless, participants (all of whom had had access to the active treatment at this point) continued to show significant improvement over baseline on all outcome variables except depression (Table 5).

Table 5. Difference between baseline and 3-month follow-up data for all participants (N=121).

Measure	*t* value *(df*=44*)*	*P* value (two-tailed)
IBS-QOL[a]	5.136	<.001
GSRS[b]	4.064	<.001
GI-COG[c]	6.090	<.001
VSI[d]	4.261	<.001
FFQ[e]	4.000	<.001
PHQ[f]	1.489	.14
DASS[g]		
Depression	0.499	.62
Stress	2.264	.03
Anxiety	3.012	.004

[a]IBS-QOL: IBS Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Finally, we assessed maintenance of treatment gains from posttreatment to 3-month follow-up. Without exception, gains were maintained, and there were no significant changes or relapse in symptoms, except for a slight rise in depression. Thus, even in the face of an incredibly stressful global pandemic, by and large, our participants showed remarkable resilience, and their HRQL, GI symptoms, GI-specific catastrophizing, anxiety, and fear of food remained much improved (Table 6). This result was confirmed in a sensitivity analysis using multiple imputation (Table 7).

Table 6. Maintenance of gains from posttreatment to 3 months.

Measure	*t* value (*df*=43)	*P* value (two-tailed)
IBS-QOL[a]	0.289	.77
GSRS[b]	0.636	.53
GI-COG[c]	0.841	.41
VSI[d]	0.056	.96
FFQ[e]	0.240	.81
PHQ[f]	–0.530	.60
DASS[g]		
Depression	–1.614	.11
Stress	0.335	.74
Anxiety	0.935	.36

[a]IBS-QOL: Irritable Bowel Syndrome-Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Table 7. Intention-to-treat sensitivity analysis of within-group changes using multiple imputation.

Measure	0-8 weeks, immediate treatment group (n=62)		8 weeks to posttreatment, waitlist group (n=59)		Posttreatment to 3-month follow-up, all participants (N=121)	
	t value	*P* value (two-tailed)	*t* value	*P* value (two-tailed)	*t* value	*P* value (two-tailed)
IBS-QOL[a]	4.7	<.001	3.5	.001	–0.47	.65
GSRS[b]	3.3	.001	2.0	.06	0.19	.85
GI-COG[c]	5.2	<.001	2.9	.008	0.10	.93
VSI[d]	3.7	<.001	2.4	.02	–0.26	.80
FFQ[e]	3.2	.002	2.0	.047	–0.15	.88
PHQ[f]	2.1	.04	2.0	.053	–0.32	.75
DASS[g]						
Depression	1.2	.23	1.8	.08	–0.87	.39
Stress	2.4	.02	1.5	.15	0.47	.64
Anxiety	2.3	.02	1.9	.07	–0.14	.89

[a]IBS-QOL: Irritable Bowel Syndrome-Quality of Life.

[b]GSRS: Gastrointestinal Symptom Rating Scale.

[c]GI-COG: Gastrointestinal Cognitions Questionnaire.

[d]VSI: Visceral Sensitivity Index.

[e]FFQ: Fear of Food Questionnaire.

[f]PHQ: Patient Health Questionnaire.

[g]DASS: Depression Anxiety Stress Scale.

Attrition

There was significant attrition from the study in both the immediate treatment and waitlist control groups (see Figure 2 for the flow chart of study enrollment). An independent-samples *t* test demonstrated that the only predictors of attrition at the 8-week follow-up were more severe visceral sensitivity (t_{119}=2.18, *P*=.03) and fear of food (t_{119}=1.79, *P*=.08) for participants in both the immediate treatment and waitlist group. About half of the participants (21 out of 44) in the waitlist control group who were offered crossover to active treatment were lost to follow-up at their posttreatment assessment. None of the measures at 8 weeks predicted attrition in this group. Of the 58 participants across both groups who completed the active treatment and the posttreatment questionnaires, 14 were lost to follow-up prior to the 3-month assessment. Participants who were lost to follow-up at that point were more likely to be *less* stressed (t_{56}=2.19, *P*=.03), catastrophized *less* (t_{56}=2.21, *P*=.03), and were somewhat *less* depressed (t_{56}=1.72, *P*=.09) at posttreatment.

Mediation

Another aim of the study was to test whether changes in catastrophic thinking, visceral sensitivity, and fear of food would at least partially mediate reductions in GI symptom severity and improvement in quality of life.

The simple mediator models for GI symptom severity showed that changes in visceral anxiety, catastrophizing, and fear of food were all significant mediators of the relationship between treatment and GI symptom severity. Participants assigned to immediate treatment had a greater decrease in visceral anxiety, catastrophizing, and fear of food, and participants who had a greater decrease in visceral anxiety, catastrophizing, and fear of food had lower GI symptom severity at 8 weeks while controlling for baseline GI symptom severity (Table 8). The statistically significant direct effect for each of the simple models indicated that treatment directly influenced GI symptom severity independent of the indirect effect of the mediating variables. The parallel multiple mediator model indicated that the indirect effects of visceral anxiety and fear of food were independent mediators, but the effect of catastrophizing was not significant (bias-corrected 95% CI included zero) and its effect is taken up by the other mediators. Once again there was a significant direct effect of treatment independent of mediators on GI symptom severity (*P*<.001).

Participants assigned to immediate treatment had a greater decrease in visceral anxiety, catastrophizing, and fear of food, and participants who had a greater decrease in visceral anxiety, catastrophizing, and fear of food had lower scores on IBS-QOL at 8 weeks while controlling for baseline IBS-QOL (Table 8). The statistically significant direct effect for the model including fear of food indicated that treatment directly influenced quality of life independent of the indirect effect of fear of food. However, having accounted for the effect of change in visceral anxiety and catastrophizing, no statistically significant direct effect of treatment remained. The parallel multiple mediator model indicated statistically significant indirect effects of the three mediators with no direct effect of treatment (Table 8).

Table 8. Direct and indirect mediation results.

Measure	GI[a] symptom severity		IBS[b] quality of life	
	Effect (95% BCI[c])	*P* value	Effect (95% BCI)	*P* value
Visceral anxiety				
Indirect	–4.3 (–7.0 to –1.8)	.002	–12.2 (–18.62 to –6.4)	<.001
Direct	–5.1 (–8.8 to –1.4)	.007	–4.6 (–9.8 to –.56)	.08
GI-specific catastrophizing				
Indirect	–3.7 (–7.1 to –1.2)	.007	–15.4 (–21.6 to –9.6)	<.001
Direct	–5.6 (–10.2 to –1.1)	.02	–1.4 (–7.9 to 5.1)	.67
Fear of food				
Indirect	–4.0 (–7.2 to –1.5)	.003	–9.8 (–16.3 to –3.8)	<.001
Direct	–5.4 (–9.1 to –1.6)	.005	–7.0 (–12.7 to –1.4)	.02
Parallel multiple mediator model				
Direct	–9.4 (–13.5 to –5.3)	<.001	–.5 (–5.2 to 4.2)	.83
Visceral anxiety	–3.4 (–6.2 to –1.0)	N/A[d]	–7.0 (–11.3 to –3.4)	N/A
COG[e]	1.8 (–1.0 to 4.4)	N/A	–5.1 (–9.0 to –1.8)	N/A
Fear of food	–2.7 (–5.5 to –0.8)	N/A	–4.3 (–8.3 to –1.2)	N/A

[a]GI: gastrointestinal.

[b]IBS: irritable bowel syndrome.

[c]BCI: bias-corrected confidence interval.

[d]N/A: not applicable.

[e]COG: cognition.

Moderation

Univariate analysis of the data revealed that Rome IV criteria moderated the effectiveness of the treatment. That is, there was a significant interaction between condition and Rome IV status such that the app was more helpful to the participants who reported meeting stringent Rome IV criteria for IBS at baseline than for those who did not, for both GI symptoms ($F_{3,76}$=2.919, *P*=*.04*<.0) and HRQL ($F_{3,76}$=6.652, *P*=.001). The only difference at baseline between those who met the criteria and those who did not was severity of GI symptoms (t_{144}=3.75, *P*<.001). No other baseline variables were significantly different. When the sample was restricted to *only* those individuals who met strict Rome IV criteria, the advantage of the treatment group over the waitlist group was even more marked for improvement in GI symptoms ($F_{1,56}$=30.2, *P*<.001), HRQL ($F_{1,56}$=47.42, *P*<.001), catastrophizing ($F_{1,56}$=51.10, *P*<.001), visceral anxiety ($F_{1,56}$=28.84, *P*<.001), and fear of food ($F_{1,56}$=22.11, *P*<.001).

We also examined whether IBS subtype moderated the efficacy of the app; it did not.

Dose-Dependent Response

Because the app itself tracks objective progress through the modules, we were able to examine the effect of "dose" (measured as components of the app accessed) on outcome. The majority of participants in the immediate treatment group who completed follow-up surveys finished just shy over 3 modules (mean 3.2, SD 2, median 2.9). Only one participant completed all possible modules. Dosage was marginally correlated with improvement in HRQL (*r*=0.33, *P*=.07) and depression (*r*=0.33, *P*=.08), but was not directly correlated with improvement in GI symptoms, changes in catastrophizing, or visceral anxiety. This suggests that the more participants used the app, the more their quality of life and depressive symptoms improved. These results also suggest that the primary change components in the app with respect to catastrophizing and visceral anxiety occurred early in the modules.

Discussion

Principal Findings

The purpose of this study was two-fold. First, we tested the efficacy of a cognitive behavioral intervention for IBS delivered via a digital self-help app, with no therapist feedback or involvement. Completer analyses yielded statistically and clinically significant improvement, with treatment having a positive impact on both GI symptom severity and quality of life. Intention-to-treat sensitivity analysis using multiple imputation replicated those findings. After treatment, individuals reported significantly lower levels of IBS symptoms and less impairment to their quality of life. Effect sizes for the primary outcomes and most of the secondary outcomes were all in the very large range. This 8-week intervention appears to have substantially reduced the burden of illness compared to that of waitlist controls.

Second, we tested whether reductions in IBS-specific catastrophic thinking, visceral sensitivity, and fear of food might mediate the efficacy of treatment. Reductions in these three variables did appear to mediate the impact of treatment on HRQL, but not on GI symptoms themselves. The app worked by reducing catastrophic thinking, visceral sensitivity to GI symptoms, and fear of food, which in turn improved individuals' quality of life. This is consistent with prior findings about the impact of CBT on IBS. Additionally, changes in catastrophizing and visceral anxiety were observed in participants who had only completed the preliminary modules of the app. This is consistent with the idea that psychoeducation and relaxation can promote cognitive reframing and can reduce anxiety about visceral sensations.

Overall, we are strongly encouraged by the results of this study, which appear to suggest that effective CBT for IBS can be successfully delivered via an app. The Zemedy app seems to be an effective means to improve the lives of individuals with IBS. Zemedy, which is already in the App Store and Google Play Store for download, could dramatically increase the accessibility of effective treatment for this debilitating disorder.

Limitations

This study had a number of limitations. The first was the lack of a placebo control condition. Patients with IBS typically show high placebo response rates [47], although the placebo effect is reduced when individuals meet more stringent (ie, Rome IV) diagnostic criteria. Future trials of the Zemedy app (including an ongoing trial registered as NCT04665271 on Clinicaltrials.gov) will include an active placebo control (sham app) rather than relying on a simple waitlist.

The second major limitation was the lack of rigorous diagnostic interviewing or explicit physician confirmation of the IBS diagnosis. Our inclusion criteria were self-reported as a prior physician diagnosis of IBS and/or meeting stringent Rome IV criteria. Five participants had no physician diagnosis but met Rome IV criteria. Thirty participants reported having been diagnosed by a physician but did not meet stringent Rome IV criteria. Of those, 7 would have met Rome III [1] criteria. The remaining individuals failed to meet either the duration or severity criteria.

The choice to include participants who did not meet Rome IV criteria was made because the aim of the study was to determine the efficacy of the app for individuals *who believe they have IBS and are searching for self-help materials*. The app will be accessible to all, and even those who perceive they have IBS without a clinical diagnosis or meeting criteria will use it. Thus, it is important to test the app among anyone who believes it to be relevant to their life. Interestingly, individuals who *did* meet criteria for IBS actually showed a significantly better response to the app. Thus, including individuals who might not have met strict Rome IV criteria is actually more conservative and more ecologically valid. The app includes educational material about the importance of a thorough differential diagnostic evaluation, and especially the importance of ruling out other potential causes of GI symptoms (such as celiac disease and inflammatory bowel diseases). Moving forward, it may be important for the app to encourage people who do *not* meet Rome IV criteria to consult with their physicians about other possible causes of their symptoms.

The third limitation was the rate of attrition, with 36% not completing follow-up measures. Of those who completed 8-week follow-up measures, most had not made it through a substantial portion of the app's content. Nevertheless, the attrition rate from treatment of 36% is actually lower than the rate of 47% on average typically found in studies of online behavioral health interventions [48].

In addition, people did not drop out entirely at random. Participants who dropped out during the initial treatment phase had significantly higher rates of visceral anxiety and fear of food at baseline (although there were no other significant differences). Since CBT for IBS typically encourages acceptance of visceral sensations and reduction of behavioral avoidance (especially avoidance of food and food-related social situations), the treatment may have seemed particularly challenging for those individuals. This might represent a population requiring more personal guidance, encouragement, and support from in-person therapy.

A fourth limitation of the study was the inability to statistically establish the temporal precedence of the proposed mediators of change. In the study design, there was no midpoint survey to show that visceral anxiety, catastrophizing, and fear of food changed *before* quality of life improved. We did not include this intermediate survey during the treatment phase because we were concerned that it would increase attrition of participants, although a future study of the app would benefit from data obtained at this point.

A fifth limitation is that the PHQ-9 is a poor measure of depression severity because it only measures symptom frequency and does not take intensity of symptoms into account. For example, at baseline, someone might indicate feeling tired or having little energy nearly every day (scoring a 3), because they are so anergic they can barely get out of bed. By the end of a trial, they might still indicate feeling tired or having little energy nearly every day (scoring a 3) because they still feel chronically fatigued, but they are getting up and going to work every day. The *severity* of their anergia would have declined significantly, but the PHQ-9 would reflect no change. Furthermore, the item that assesses suicidality makes no distinctions at all with respect to passive versus active ideation, nor does it capture intent. An individual who has passive suicidal ideation daily, but no intent, would actually score *higher* than an individual who experiences less frequent, but intense active suicidal ideation with wavering intent. Although the PHQ-9 has been used in many other clinical trials of behavioral health interventions, and it *did* show significant improvement over the course of this trial in the completer sample (but not in the intention-to-treat analyses), we were dissatisfied with its sensitivity to treatment effects. Future studies of the app will employ more sensitive measures.

A sixth limitation is that we did not assess concurrent medication use. However, there is no reason to believe that medication use would have been different across the immediate treatment and control groups.

Finally, the last phase of the trial occurred during the COVID-19 global pandemic. Since all waitlist participants had already been crossed over to the active treatment phase, the 3-month follow-up data may be less reflective of the enduring effects of the treatment and more reflective of the massive social, economic, and personal upheaval the pandemic has caused. Indeed, the end of the treatment phase for all participants coincided with the COVID-19 pandemic's arrival in the United States. With massive shutdowns and quarantines, it is highly likely that distress increased for all participants. The fact that treatment gains were generally maintained and that participants remained much improved over baseline (except for some recurrence of depression), even in the face of an unprecedented global health crisis, is encouraging.

Conclusion

Despite the limitations, we believe that this study is of significant value. It successfully demonstrated the efficacy of an app that provided CBT for IBS patients. The intervention was not restricted by geography or scheduling constraints, and required no face-to-face contact with a clinician, aspects that dramatically increase the accessibility and portability of treatment. Despite its relatively benign physical profile, IBS can be an extraordinarily debilitating condition. Finding novel ways to disseminate evidence-based, effective treatments remains an important challenge, and Zemedy is a promising and effective way to help those suffering from IBS.

Acknowledgments

Bold Health provided some funding to pay for recruitment advertisements and participant incentives.

Authors' Contributions

MH substantively designed the content of the intervention, designed and ran the trial, performed most of the statistical analyses, and wrote most of the paper. SM and BD were the research coordinators who oversaw much of the day-to-day administration of the trial, including recruitment, randomization, scheduling assessments, and scoring and cleaning the data. In addition, SM assisted with statistical analyses and drafting the manuscript. OO oversaw the conversion of the CBT protocol into the digital format, coordinating the team of programmers and designers who created the app itself. SW performed the intention-to-treat and mediation analyses, and drafted those sections of the manuscript. All authors reviewed and approved the manuscript.

Conflicts of Interest

OO has a financial ownership stake in Bold Health, which developed and markets the Zemedy app. MH, SM, BD, and SW have no conflicts to declare.

Multimedia Appendix 1
Consent form.
[DOCX File , 16 KB - mhealth_v9i5e26152_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (v 1.6.1).
[PDF File (Adobe PDF File), 9268 KB - mhealth_v9i5e26152_app2.pdf]

References

1. Black CJ, Yiannakou Y, Houghton LA, Ford AC. Epidemiological, clinical, and psychological characteristics of individuals with self-reported irritable bowel syndrome based on the Rome IV vs Rome III criteria. Clin Gastroenterol Hepatol 2020 Feb;18(2):392-398. [doi: 10.1016/j.cgh.2019.05.037] [Medline: 31154027]
2. Yeh H, Chien W, Chung C, Hu J, Tzeng N. Risk of psychiatric disorders in irritable bowel syndrome-A nationwide, population-based, cohort study. Int J Clin Pract 2018 Jul 19;72(7):e13212. [doi: 10.1111/ijcp.13212] [Medline: 29920876]
3. Hunt M. Cognitive-behavioral therapy for irritable bowel syndrome. In: Sobin W, editor. Using central neuromodulators and psychological therapies to manage patients with disorders of gut-brain interaction. Switzerland: Springer, Cham; 2019:95-141.
4. Simrén M, Törnblom H, Palsson OS, Van Oudenhove L, Whitehead WE, Tack J. Cumulative effects of psychologic distress, visceral hypersensitivity, and abnormal transit on patient-reported outcomes in irritable bowel syndrome. Gastroenterology 2019 Aug;157(2):391-402. [doi: 10.1053/j.gastro.2019.04.019] [Medline: 31022401]
5. Zhang Y, Qin G, Liu D, Wang Y, Yao S. Increased expression of brain-derived neurotrophic factor is correlated with visceral hypersensitivity in patients with diarrhea-predominant irritable bowel syndrome. World J Gastroenterol 2019 Jan 14;25(2):269-281 [FREE Full text] [doi: 10.3748/wjg.v25.i2.269] [Medline: 30670915]
6. Labus JS, Mayer EA, Chang L, Bolus R, Naliboff BD. The central role of gastrointestinal-specific anxiety in irritable bowel syndrome: further validation of the visceral sensitivity index. Psychosom Med 2007 Jan;69(1):89-98. [doi: 10.1097/PSY.0b013e31802e2f24] [Medline: 17244851]

7. Addante R, Naliboff B, Shih W, Presson AP, Tillisch K, Mayer EA, et al. Predictors of health-related quality of life in irritable bowel syndrome patients compared with healthy individuals. J Clin Gastroenterol 2019 Apr;53(4):e142-e149 [FREE Full text] [doi: 10.1097/MCG.0000000000000978] [Medline: 29351154]
8. Sherwin L, Leary E, Henderson WA. The association of catastrophizing with quality-of-life outcomes in patients with irritable bowel syndrome. Qual Life Res 2017 Aug;26(8):2161-2170 [FREE Full text] [doi: 10.1007/s11136-017-1554-0] [Medline: 28324323]
9. Sugaya N, Kaiya H, Kumano H, Nomura S. Relationship between subtypes of irritable bowel syndrome and severity of symptoms associated with panic disorder. Scand J Gastroenterol 2008;43(6):675-681. [doi: 10.1080/00365520701883478] [Medline: 18569984]
10. Kinsinger S. Cognitive-behavioral therapy for patients with irritable bowel syndrome: current insights. Psychol Res Behav Manag 2017;10:231-237. [doi: 10.2147/PRBM.S120817] [Medline: 28790872]
11. Radziwon CD, Lackner JM. Cognitive behavioral therapy for IBS: how useful, how often, and how does it work? Curr Gastroenterol Rep 2017 Aug 17;19(10):49-49. [doi: 10.1007/s11894-017-0590-9] [Medline: 28819814]
12. Henrich JF, Gjelsvik B, Surawy C, Evans E, Martin M. A randomized clinical trial of mindfulness-based cognitive therapy for women with irritable bowel syndrome-Effects and mechanisms. J Consult Clin Psychol 2020 Apr;88(4):295-310. [doi: 10.1037/ccp0000483] [Medline: 32134291]
13. Laird K, Tanner-Smith E, Russell A, Hollon S, Walker L. Short-term and long-term efficacy of psychological therapies for irritable bowel syndrome: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2016 Jul;14(7):937-947 [FREE Full text] [doi: 10.1016/j.cgh.2015.11.020] [Medline: 26721342]
14. Shah K, Ramos-Garcia M, Bhavsar J, Lehrer P. Mind-body treatments of irritable bowel syndrome symptoms: An updated meta-analysis. Behav Res Ther 2020 May;128:103462. [doi: 10.1016/j.brat.2019.103462] [Medline: 32229334]
15. Hunt MG, Moshier S, Milonova M. Brief cognitive-behavioral internet therapy for irritable bowel syndrome. Behav Res Ther 2009 Sep;47(9):797-802. [doi: 10.1016/j.brat.2009.05.002] [Medline: 19570525]
16. Craske MG, Wolitzky-Taylor KB, Labus J, Wu S, Frese M, Mayer EA, et al. A cognitive-behavioral treatment for irritable bowel syndrome using interoceptive exposure to visceral sensations. Behav Res Ther 2011 Jun;49(6-7):413-421 [FREE Full text] [doi: 10.1016/j.brat.2011.04.001] [Medline: 21565328]
17. Windgassen S, Moss-Morris R, Chilcot J, Sibelli A, Goldsmith K, Chalder T. The journey between brain and gut: A systematic review of psychological mechanisms of treatment effect in irritable bowel syndrome. Br J Health Psychol 2017 Nov 01;22(4):701-736. [doi: 10.1111/bjhp.12250] [Medline: 28573818]
18. Inadomi J, Fennerty MB, Bjorkman D. Systematic review: the economic impact of irritable bowel syndrome. Aliment Pharmacol Ther 2003 Oct 01;18(7):671-682 [FREE Full text] [doi: 10.1046/j.1365-2036.2003.t01-1-01736.x] [Medline: 14510740]
19. Ljótsson B, Hesser H, Andersson E, Lackner JM, El Alaoui S, Falk L, et al. Provoking symptoms to relieve symptoms: a randomized controlled dismantling study of exposure therapy in irritable bowel syndrome. Behav Res Ther 2014 Apr;55:27-39. [doi: 10.1016/j.brat.2014.01.007] [Medline: 24584055]
20. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, ACTIB trial group. Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. Gut 2019 Sep;68(9):1613-1623 [FREE Full text] [doi: 10.1136/gutjnl-2018-317805] [Medline: 30971419]
21. Hunt M. Reclaim your life from IBS: A scientifically proven plan for relief without restrictive diets. New York: Sterling; Apr 05, 2016.
22. Hunt MG, Ertel E, Coello JA, Rodriguez L. Empirical support for a self-help treatment for IBS. Cogn Ther Res 2014 Oct 19;39(2):215-227. [doi: 10.1007/s10608-014-9647-3]
23. Sumant O. mHealth Market Size, Share and Trends: Industry Growth, 2027. Allied Market Research. 2020. URL: https://www.alliedmarketresearch.com/mobile-health-market [accessed 2020-04-20]
24. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
25. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
26. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a Web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013 Sep 02;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
27. Kelders SM, Bohlmeijer ET, Pots WTM, van Gemert-Pijnen JEWC. Comparing human and automated support for depression: Fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]
28. Mason EC, Andrews G. The use of automated assessments in internet-based CBT: The computer will be with you shortly. Internet Interv 2014 Oct;1(4):216-224. [doi: 10.1016/j.invent.2014.10.003]

29. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. Drossman DA, Patrick DL, Whitehead WE, Toner BB, Diamant NE, Hu Y, et al. Further validation of the IBS-QOL: a disease-specific quality-of-life questionnaire. Am J Gastroenterol 2000 Apr;95(4):999-1007. [doi: 10.1111/j.1572-0241.2000.01941.x] [Medline: 10763950]
32. Patrick DL, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998 Feb;43(2):400-411. [doi: 10.1023/a:1018831127942] [Medline: 9512138]
33. Wiklund IK, Fullerton S, Hawkey CJ, Jones RH, Longstreth GF, Mayer EA, et al. An irritable bowel syndrome-specific symptom questionnaire: development and validation. Scand J Gastroenterol 2003 Sep;38(9):947-954. [doi: 10.1080/00365520310004209] [Medline: 14531531]
34. Drossman DA, Chang L, Bellamy N, Gallo-Torres HE, Lembo A, Mearin F, et al. Severity in irritable bowel syndrome: a Rome Foundation Working Team report. Am J Gastroenterol 2011 Oct;106(10):1749-59; quiz 1760. [doi: 10.1038/ajg.2011.201] [Medline: 21747417]
35. Palsson OS, van Tilburg MA, Simren M, Sperber AD, Whitehead WE. Mo1642 Population prevalence of Rome IV and Rome III irritable bowel syndrome (IBS) in the United States (US), Canada and the United Kingdom (UK). Gastroenterology 2016 Apr;150(4):S739-S740. [doi: 10.1016/s0016-5085(16)32513-6]
36. Hunt M, Zickgraf H, Gibbons B, Loftus P. Development and validation of the Fear of Food Questionnaire (FFQ). 2018 Presented at: Annual Meeting of the Anxiety and Depression Association of America; April 19, 2018; Washington, DC.
37. Labus J, Bolus R, Chang L, Wiklund I, Naesdal J, Mayer EA, et al. The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. Aliment Pharmacol Ther 2004 Jul 01;20(1):89-97. [doi: 10.1111/j.1365-2036.2004.02007.x] [Medline: 15225175]
38. Hazlett-Stevens H, Craske MG, Mayer EA, Chang L, Naliboff BD. Prevalence of irritable bowel syndrome among university students. J Psychosom Res 2003 Dec;55(6):501-505. [doi: 10.1016/s0022-3999(03)00019-9]
39. Hunt M, Ertel E, Coello JA, Rodriguez L. Development and validation of the GI-Cognitions Questionnaire. Cogn Ther Res 2014 Mar 4;38(4):472-482. [doi: 10.1007/s10608-014-9607-y]
40. Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behav Res Ther 1997 Jan;35(1):79-89. [doi: 10.1016/s0005-7967(96)00068-x]
41. Parkitny L, McAuley J. The Depression Anxiety Stress Scale (DASS). J Physiother 2010;56(3):204. [doi: 10.1016/s1836-9553(10)70030-8]
42. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003 Jun;42(Pt 2):111-131. [doi: 10.1348/014466503321903544] [Medline: 12828802]
43. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
44. Nevalainen J, Kenward MG, Virtanen SM. Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. Stat Med 2009 Dec 20;28(29):3657-3669. [doi: 10.1002/sim.3731] [Medline: 19757484]
45. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach, Second edition. New York: Guilford Press; Dec 13, 2017.
46. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
47. Patel SM, Stason WB, Legedza A, Ock SM, Kaptchuk TJ, Conboy L, et al. The placebo effect in irritable bowel syndrome trials: a meta-analysis. Neurogastroenterol Motil 2005 Jun;17(3):332-340. [doi: 10.1111/j.1365-2982.2005.00650.x] [Medline: 15916620]
48. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 01;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]

Abbreviations

CBT: cognitive behavioral therapy
DASS: Depression Anxiety Stress Scale
FFQ: Fear of Food Questionnaire
GI: gastrointestinal
GI-COG: Gastrointestinal Cognitions Questionnaire
GSRS: Gastrointestinal Symptom Rating Scale
HRQL: health-related quality of life

IBS: irritable bowel syndrome
IBS-QOL: IBS Quality of Life Impairment
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
VSI: Visceral Sensitivity Index

Edited by G Eysenbach; submitted 04.12.20; peer-reviewed by S Weinland, S Mathews; comments to author 29.12.20; revised version received 09.02.21; accepted 14.04.21; published 20.05.21.

Please cite as:
Hunt M, Miguez S, Dukas B, Onwude O, White S
Efficacy of Zemedy, a Mobile Digital Therapeutic for the Self-management of Irritable Bowel Syndrome: Crossover Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e26152
URL: https://mhealth.jmir.org/2021/5/e26152
doi:10.2196/26152
PMID:33872182

Original Paper

Comparative Analysis of Single and Combined Antipyretics Using Patient-Generated Health Data: Retrospective Observational Study

Yu Rang Park[1*], PhD; Hyery Kim[2*], MD, PhD; Ji Ae Park[1], MS; Sang Hyun Ahn[3], MD; Seyun Chang[4], BS; Jae Won Shin[4], MD; Myeongchan Kim[4], BS; Jae-Ho Lee[5,6], MD, PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Pediatrics, University of Ulsan College of Medicine, Asan Medical Center Children's Hospital, Seoul, Republic of Korea

[3]Korea Human Resource Development Institute for Health & Welfare, Cheongju, Republic of Korea

[4]Mobile Doctor Co, Ltd, Seoul, Republic of Korea

[5]Department of Information Medicine, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Republic of Korea

[6]Department of Emergency Medicine, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jae-Ho Lee, MD, PhD
Department of Information Medicine
Asan Medical Center
College of Medicine, University of Ulsan
88 Olympic-ro 43 gil
Songpa gu
Seoul
Republic of Korea
Phone: 82 0230104185
Email: rufiji@gmail.com

Abstract

Background: Fever is one of the most common symptoms in children and is the physiological response of the human immune system to external pathogens. However, effectiveness studies of single and combined antipyretic therapy are relatively few due to lack of data. In this study, we used large-scale patient-generated health data from mobile apps to compare antipyretic affects between single and combination antipyretics.

Objective: We aimed to establish combination patterns of antipyretics and compare antipyretic affects between single and combination antipyretics using large-scale patient-generated health data from mobile apps.

Methods: This study was conducted using medical records of feverish children from July 2015 to June 2017 using the Fever Coach mobile app. In total, 3,584,748 temperature records and 1,076,002 antipyretic records of 104,337 children were analyzed. Antipyretic efficacy was measured by the mean difference in the area under the temperature change curve from baseline for 6 hours, 8 hours, 10 hours, and 12 hours after antipyretic administration in children with a body temperature of ≥38.0 ˚C between single and combination groups.

Results: The single antipyretic and combination groups comprised 152,017 and 54,842 cases, respectively. Acetaminophen was the most commonly used single agent (60,929/152,017, 40.08%), and acetaminophen plus dexibuprofen was the most common combination (28,065/54,842, 51.17%). We observed inappropriate use, including triple combination (1205/206,859, 0.58%) and use under 38 ˚C (11,361/206,859, 5.50%). Combination antipyretic use increased with temperature; 23.82% (33,379/140,160) of cases were given a combination treatment when 38 ˚C ≤ temperature < 39 ˚C, while 41.40% (1517/3664) were given a combination treatment when 40 ˚C ≤ temperature. The absolute value of the area under the curve at each hour was significantly higher in the single group than in the combination group; this trend was consistently observed, regardless of the type of antipyretics. In particular, the delta fever during the first 6 hours between the two groups showed the highest difference. The combination showed the lowest delta fever among all cases.

Conclusions: Antipyretics combination patterns were analyzed using large-scale data. Approximately 75% of febrile cases used single antipyretics, mostly acetaminophen, but combination usage became more frequent as temperature increased. However, combination antipyretics did not show definite advantages over single antipyretics in defervescence, regardless of the combination. Single antipyretics are effective in reducing fever and relieving discomfort in febrile children.

(JMIR Mhealth Uhealth 2021;9(5):e21668) doi:10.2196/21668

KEYWORDS

combination antipyretics; fever management; patient-generated health data; comparative analysis; mHealth; apps; fever

Introduction

Fever is one of the most common symptoms in children and represents a physiologic response of the human immune system against external pathogens. In particular, fever is reported to be the most common cause of children's emergency room visits, accounting for 26.4% to 37.4% of visits [1,2]. According to studies in Korea, most causes of childhood fever are viral infections such as mild acute upper respiratory infections [1,2]. However, as parents wish to lower their child's temperature to within a normal range, antipyretics are commonly administered [3]. This fever phobia has caused caregivers to treat fever aggressively with a combination of antipyretics such as acetaminophen (ACE) and ibuprofen (IBU), often in combination. Despite the lack of official recommendations, combination antipyretic therapy with ACE and IBU is commonly used to treat fever that does not respond to monotherapy [4].

Pharmacologic evidence suggests that the combination of ACE and IBU may be well tolerated because both medications have different metabolic pathways that are not affected by each other [5]. Both drugs are well tolerated, with wide therapeutic margins and proper dosing.

However, because each antipyretic has its own adverse effects, the combination of antipyretics should be properly guided based on reference data regarding antipyretic efficacy. Adverse effects caused by ACE and IBU are well known because of their frequent use worldwide. Hepatic disorders are the most common adverse drug reaction of ACE, whereas gastrointestinal tract disorder, kidney injury, hypersensitivity reaction, and, more recently, noticeable cardiovascular events are frequently reported adverse effects of nonsteroidal anti-inflammatory drugs (NSAIDs) such as IBU [6,7]. In addition, drug misdose or overdose is an important public health issue, especially with antipyretics because most antipyretics are administered to children. Misdosing of antipyretics has been reported in approximately 50% of ACE and 25% of IBU administrations [8-10]; among these, 8.4% to 14% and 9.6% to 14% of patients received higher doses, respectively. Infants aged younger than 1 year were more likely to receive an inaccurate dose than older children because caregivers who used inaccurate doses had difficulty in dosing based on their children's weight [11]. Overdose of ACE is one of the leading causes of acute liver failure in Western countries [10]. In addition, elevated liver transaminases have been described even at recommended doses in children as well as adults [12].

Several randomized controlled trials that have incorporated combination antipyretics have compared the antipyretic effect and toxicity; these studies favored combination therapy for achieving and sustaining an afebrile state [5,10,13-15]. However, the trial limitations included relatively small numbers of patients and a lack of continuously tracked temperatures, especially in the home setting where combination therapy was administered. Thus, concerns regarding whether combination antipyretics have substantial benefit remain.

To address this concern, drug administration and corresponding temperature recordings are required in a large cohort. These data can be obtained through mobile apps where caregivers can record their child's data in real time. In this study, we used the mobile app Fever Coach (Mobile Doctor) to obtain real-world data to determine the combination antipyretic use pattern and effects administered at home in numerous febrile children in Korea and compared the antipyretic efficacy between single and combination antipyretics.

Methods

Design

To compare the fever reduction effect between single and combination antipyretics, we used temperature and antipyretic records from feverish children collected from July 2015 to June 2017 using the Fever Coach app.

We defined the onset of a fever case when the temperature exceeded 38.0 ˚C, and the offset of a fever case when the temperature fell below 38.0 ˚C [16,17]. Temperature values were obtained using linear imputation techniques when values were missing between two neighboring actual temperatures. We used this linear imputation technique with the assumption that fever progression would show a linear characteristic between two short measurements.

The duration was defined as the time elapsed between onset and offset. Antipyretic efficacy was measured by the mean difference between the area under the curve (AUC) and the average temperature changes of the single and combination groups. Because both the degree of temperature elevation and the duration of fever affect patients, the AUC (the product of body temperature and duration) has been used as an indicator for total fever exposure in children and was the study end point to determine the effect of drugs including antipyretics [18-20]. The AUC was calculated using the area under the temperature change curve from baseline for 6 hours, 8 hours, 10 hours, and 12 hours after antipyretic administration in children with a body temperature of ≥38.0 ˚C. The populations of single and combination groups were then analyzed with two threshold

temperatures at the time of first antipyretic administration: >38 or 39 ˚C. We compared the mean difference of the negative AUC from onset through 6 hours, 8 hours, 10 hours, and 12 hours and the average temperature changes over time according to the antipyretic administration pattern.

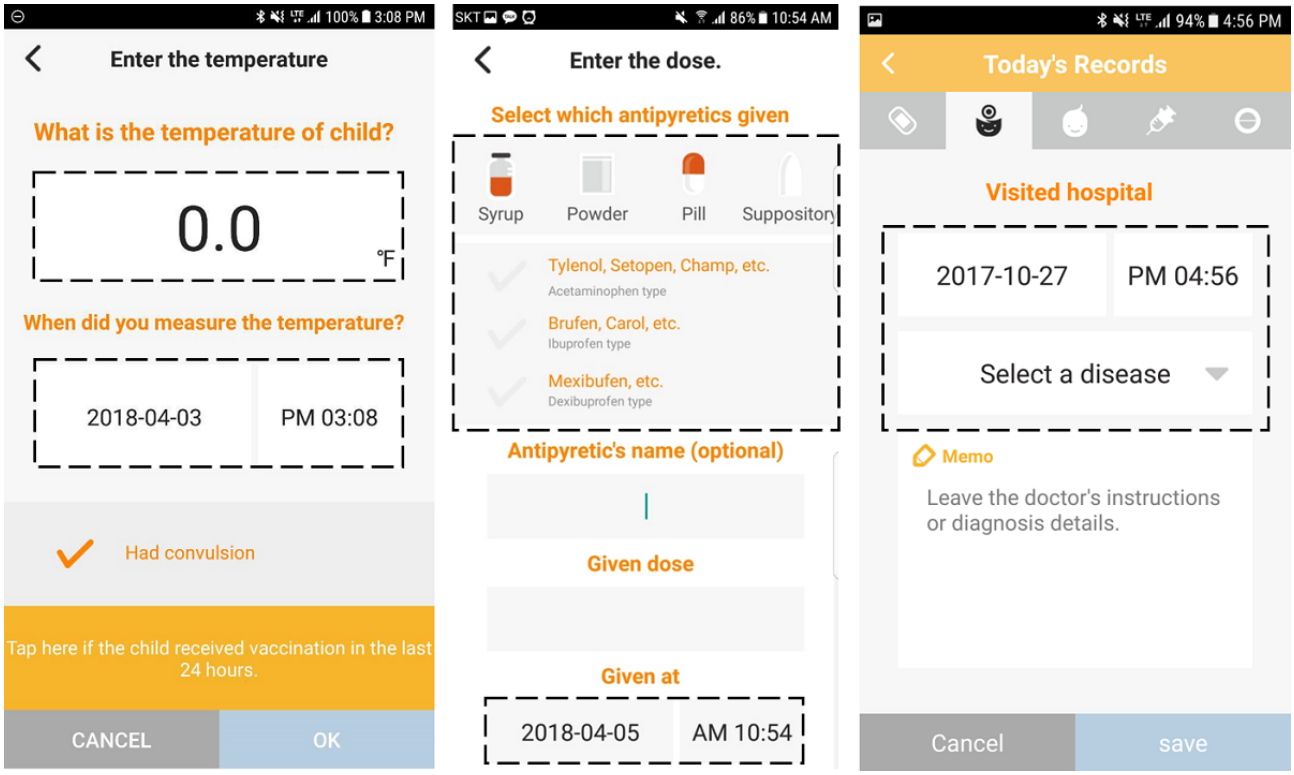
This study was approved by the institutional review board of Asan Medical Center (2018-0179). The need for informed consent was waived by the ethics committee as this study used routinely collected log data that were anonymously managed at all stages, including during data cleaning and statistical analyses. This study was conducted in accordance with the STROBE statement (Multimedia Appendix 1).

Setting

The app is based on pediatric thermal standards and assesses a child's condition based on user input to provide guidelines for antipyretic use. After registration, the user may enter various fever-related information including body temperature, other symptoms, dose and time of antipyretic administration, vaccination and antibiotic history, and physician-made diagnosis if the child was seen by a physician for the management of the current illness. Based on these inputs, the app provides instructions for fever control and advises on the timing of seeking medical attention. The service was designed and reviewed by two board-certified family physicians and one board-certified pediatrician [21]. The app supports the effective and accurate control of common fever symptoms and is available as a free download from Google Play and the App Store. As of June 31, 2017, 393,700 users had registered their child with the app.

The items in the dashed boxes show the detailed attributes on temperature and antipyretic records and the diagnosis of feverish children (Figure 1). The "Enter the temperature" function records the temperature, time, and experience of febrile convolutions. The "Enter the dose" function records the antipyretic type, ingredients, brand name, doses, and time. The "Today's records" function records the date and time of the hospital visit, diagnosis, and the doctor's instruction for the child. The diagnosis includes 21 febrile illnesses that are common in children and diseases that are directly entered.

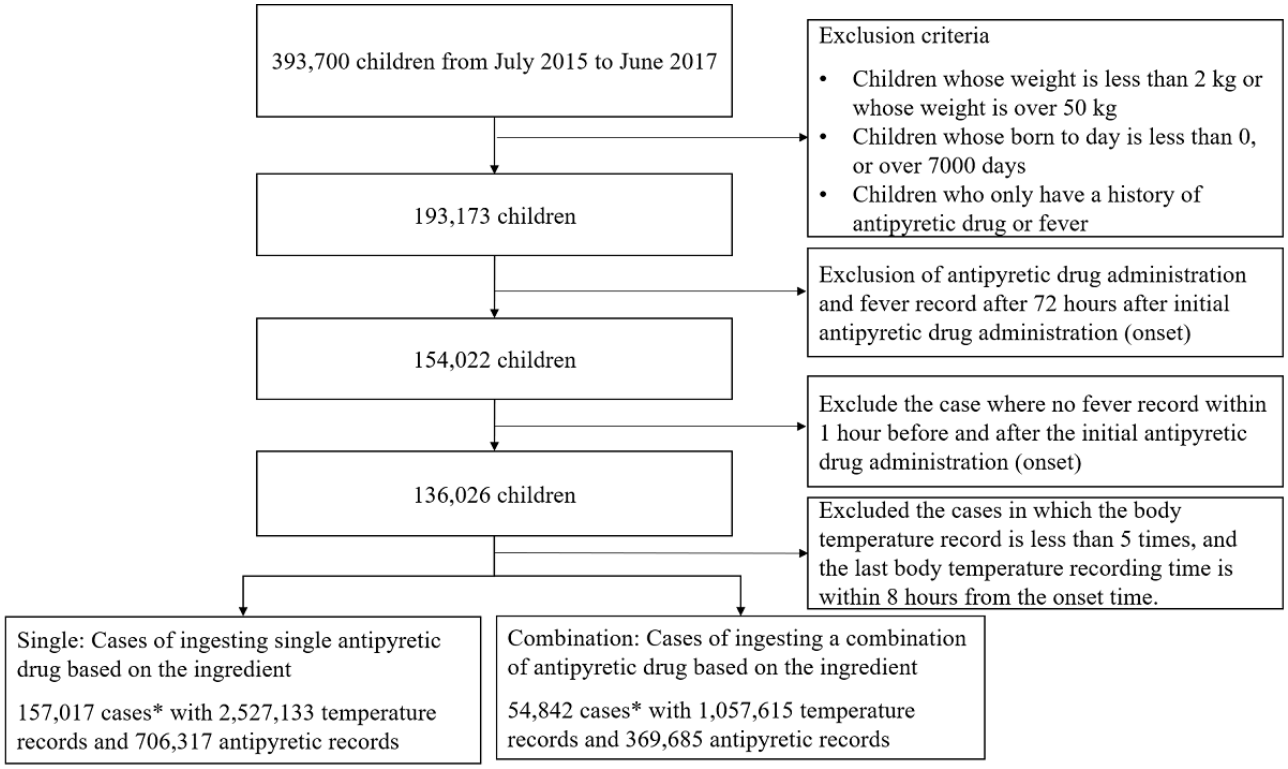
Figure 1. Antipyretic and temperature screens in the Fever Coach app: Enter the temperature (left), Enter the dose (center), Enter the diagnosis (right). The original app was in Korean.



Sample and Sampling

We analyzed the log of all users and their children who signed up and logged in to the Fever Coach app more than once from July 2015 to June 2017. Figure 2 shows the target population selection flow of the study. A total of 5,580,762 body temperatures and 1,693,295 antipyretic records were collected on 393,700 children. Among them, we excluded children weighing <2 kg or >50 kg, children <0 days or >7000 days after birth, and children without antipyretic or fever records.

Figure 2. Data collection flow chart. One case refers to a single record for 72 hours after the first antipyretic drug is administered.



Since a child may experience multiple fevers, one case indicated a single record for 72 hours after the first antipyretic drug was administered. We obtained 494,929 cases with 4,891,960 temperature records and 1,666,346 antipyretic records.

For 72 hours, cases with only one type of antipyretic were defined as a single group (152,019 cases with 2,527,133 temperature and 706,317 antipyretic records), and those including more than one type of antipyretic were recorded as a combination group (54,842 cases with 1,057,615 temperature and 369,685 antipyretic records).

Statistical Analysis

We compared the demographic and baseline characteristics of children between single and combination groups by means and frequencies using the Student *t* test and chi-square test, respectively. We also compared demographic and body temperature information according to antipyretic administration patterns in single and combination groups by Student *t* test and chi-square test. For the antipyretic effect of single and combination treatment, the AUC values and the fever reduction indicators were analyzed by groups using Student *t* test based on the temperature at the first antipyretic administration (>38 ˚C group, >39 ˚C group). We analyzed the average body temperature change over time by single versus combination and each type of antipyretic administration using a line graph. The effect sizes were calculated using Cohen *d* coefficient interpretation [22]. All reported *P* values were 2-sided, and *P*<.05 was considered significant. Data analyses were conducted using R software (version 4.0.3, R Foundation for Statistical Computing).

Results

Overall Characteristics

From July 2015 to June 2017, 3,584,748 temperature records and 1,076,002 antipyretic medication records for 104,337 children were analyzed. Among 206,859 fever cases, the single group comprised 152,017 cases, which encompassed 2,527,133 temperature and 706,317 antipyretic records (Table 1). The combination group comprised 54,842 cases, which encompassed 1,057,615 temperature and 369,685 antipyretic records. When antipyretics were used in combination, they were mostly administered alternatively at 4- to 6-hour intervals (data not shown). As shown in Table 1, most variables were statistically different between the single and combination groups, except the presence of a diagnosis. The proportion of males was slightly higher in the combination group, but no definite male or female predominance was found in either group. Children in the single group were younger (mean 830.49 days) than those in the combination group (mean 848.83 days). The mean temperature at first antipyretic administration was 38.63 ˚C in all children and was higher, at 38.76 ˚C, in the combination group and lower, at 38.59 ˚C, in the single group, with statistical significance (*P*<.001). The mean maximum temperature in the combination group was higher than that in the single group (single 38.98 ˚C vs combination 39.29 ˚C; *P*<.001). However, the mean fever duration was shorter in the single group than in the combination (single 25.08 hours vs combination 28.81 hours; *P*<.001).

When cases were divided by the onset temperature, the most common onset temperatures were between 38 and 39 ˚C in 140,160 cases, followed by those between 39 and 40 ˚C in 51,674 cases. There were 5.49% (11,361/206,859) of cases in whom antipyretics were administered under 38 ˚C. In particular,

0.18% (377/206,859) of cases used antipyretics under 37 ˚C (Table 1). The use of combination antipyretics increased with temperature: 23.81% (33,379/140,160) used a combination when the temperature was between 38 and 39 ˚C, while 41.40% (1517/3664) used a combination when the temperature was >40 ˚C. In contrast, use of single antipyretics decreased from 76.19% (106,781/140,160) to 58.60% (2147/3664) as the temperature increased.

Table 1. Descriptive statistics and children's temperature records according to antipyretic drug administration pattern (single or combination).

Variable	Single (n=152,017)	Combination (n=54,842)	*P* value[a]	Effect size[b]	Total (n=206,859)
Sex, male, n (%)	76,164 (50.10)	28,173 (51.37)	<.001	0.025	104,337 (50.44)
Age (days), mean (SD)	830.49 (534.18)	848.83 (517.23)	<.001	0.035	835.35 (529.80)
Temperature records, n (%)	2,527,133 (70.50)	1,057,615 (29.50)	—[c]	—	3,584,748
Antipyretic records, n (%)	706,317 (65.64)	369,685 (34.36)	—	—	1,076,002
Temperature at first antipyretic,˚C administration,˚C, mean (SD)	38.59 (0.53)	38.76 (0.56)	<.001	0.31	38.63 (0.54)
Maximum temperature, ˚C, mean (SD)	38.98 (0.67)	39.29 (0.58)	<.001	0.49	39.09 (0.65)
Diagnosis[d] (ICD-10 CM code), n (%)	1229 (0.81)	457 (0.83)	.58	0.002	1686 (0.82)
Acute upper respiratory infections of multiple and unspecified sites/acute tonsillitis (J06, J03), n (%)	244 (19.985)	105 (22.98)	.16	0.41	349 (20.70)
Acute nasopharyngitis (J00), n (%)	215 (17.49)	76 (16.63)	.68	0.38	291 (17.26)
Influenza and pneumonia (J09-J11), n (%)	193 (15.70)	73 (15.98)	.89	0.37	266 (15.77)
Duration of fever, hours, mean (SD)	25.08 (17.38)	28.81 (16.40)	<.001	0.22	26.02 (17.22)
Cases according to the onset temperature, n (%)					
Temperature < 37 ˚C	328 (87.00)	49 (13.00)	<.001	0.03	377
37 ˚C ≤ Temperature < 38 ˚C	8905 (81.07)	2079 (18.93)	<.001	0.09	10,984
38 ˚C ≤ Temperature < 39 ˚C	106,781 (76.19)	33,379 (23.81)	<.001	0.19	140,160
39 ˚C ≤ Temperature < 40 ˚C	33,856 (65.52)	17,818 (34.48)	<.001	0.23	51,674
Temperature ≥ 40 ˚C	2147 (58.60)	1517 (41.40)	<.001	0.09	3664

[a]Student *t* test between the single and combination groups.

[b]Cohen effect size: 0.2, small; 0.5, medium; 0.8, high.

[c]Not applicable.

[d]Diagnosis included within 1 week before and after onset of fever. The 3 most frequently registered diagnoses are listed in order of frequency.

Antipyretic Drug Administration Pattern

Within the single group, ACE was the most commonly administered (60,929/152,017, 40.08%) single agent, followed by dexibuprofen (DEX) in 36.74% (55,847/152,017) and IBU in 23.18% (35,241/152,017; Multimedia Appendix 2). Within the single group, the mean age was lower in children who were administered ACE than those who were administered IBU or DEX (ACE: 714.27 days, IBU: 902.99 days, DEX: 911.54 days).

In the combination group, the ACE-DEX combination comprised more than half of the total (28,065/54,842, 51.17%), followed by ACE-IBU (22,277/54,842, 40.62%) and IBU-DEX (3295/54,842, 6.01%). Ingesting three antipyretics (ACE-IBU-DEX) accounted for 2.20% (1205/54,842).

In the single group, ACE was the most administered drug when the temperature was between 37 and 38 ˚C in 39.66% (4356/10,984) of cases and between 38 and 39 ˚C in 30.99% (43,442/140,160) of cases. DEX was the second most administered antipyretic agent when the onset temperature was <40 ˚C but was the most administered when the temperature was >40 ˚C in 23.36% (856/3664) of cases. In the combination group, ACE-DEX was the most used combination, regardless of temperature.

ACE as a single agent was not the preferred choice at higher temperatures (43,442/140,160, 30.99%, vs 12,188/51,674, 23.59%, vs 750/3664, 20.47%: 38 to 39 ˚C vs 39 to 40 ˚C vs ≥40 ˚C); however, IBU-DEX was used at a similar frequency, regardless of temperature (2032/140,160, 1.45%, vs 1041/51,674, 2.01%, vs 80/3664, 2.18%: 38 to 39 ˚C vs 39 to 40 ˚C vs ≥40 ˚C).

Comparison of Efficacy Between Different Patterns of Antipyretics

For the antipyretic effect of single and combination treatment, the changes of AUC values, as the indicator for total fever exposure, were analyzed by groups based on the temperature at the first antipyretic administration (>38 ˚C group, >39 ˚C group; Table 2). The results according to antipyretic ingredients

are summarized in Multimedia Appendix 3. The single group showed significantly higher absolute AUC values than the combination group. In particular, the difference in the AUC during the first 6 hours of fever showed the highest effect size than other times. In the >38 ˚C group, IBU showed the highest AUC at all times, and ACE-IBU showed the highest AUC except for the first 6 hours among combinations. Moreover, the AUC at 12 hours was higher with IBU (–8.00 [SD 5.70] than with ACE-IBU –6.78 [SD 5.21]). Similarly, in the group >39 ˚C, IBU had the highest AUC value, and ACE-IBU had the highest value among combinations. Triple combination antipyretics with ACE + IBU + DEX showed the lowest AUC.

Table 2. Comparison of the area under the curve between administration patterns.

Time from the onset of fever (hours)	Single, AUC[a] (SD)	Combination, AUC (SD)	*P* value[b], AUC (SD)	Effect size[c], AUC (SD)	Total, AUC (SD)
Temperature above 38 ˚C at the time of first administration of antipyretic					
6	−3.99 (2.87)	−2.86 (2.39)	<.001	0.42	−3.69 (2.79)
8	−5.21 (3.79)	−4.15 (3.3)	<.001	0.29	−4.92 (3.7)
10	−6.42 (4.62)	−5.41 (4.21)	<.001	0.22	−6.15 (4.54)
12	−7.68 (5.47)	−6.64 (5.13)	<.001	0.19	−7.4 (5.4)
Temperature above 39 ˚C at the time of first antipyretic					
6	−5.54 (3.29)	−4.31 (2.64)	<.001	0.41	−5.11 (3.14)
8	−7.4 (4.33)	−6.23 (3.61)	<.001	0.29	−6.99 (4.13)
10	−9.31 (5.27)	−8.11 (4.58)	<.001	0.24	−8.89 (5.07)
12	−11.33 (6.24)	−10 (5.55)	<.001	0.22	−10.8 6 (6.04)

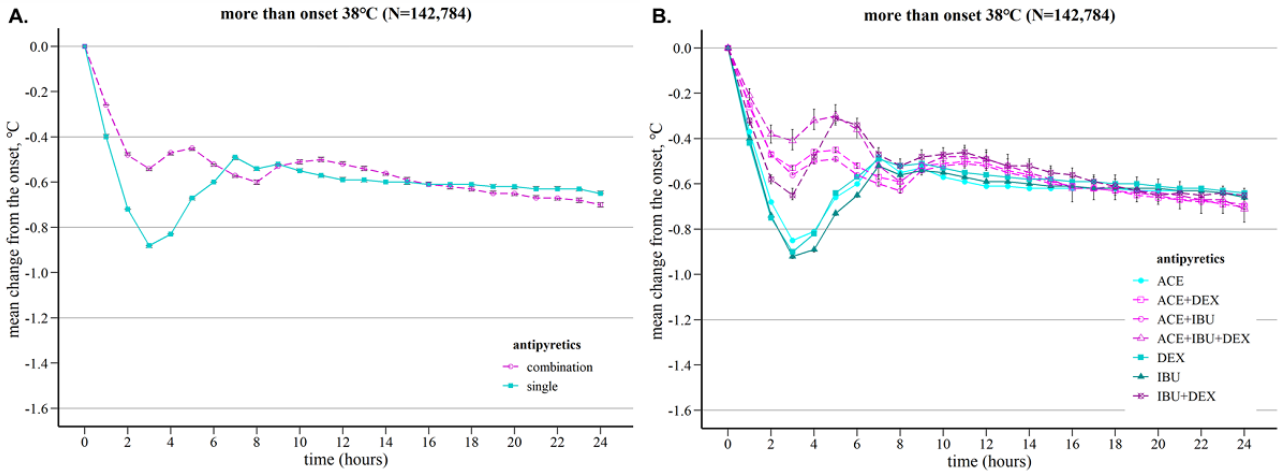
[b]AUC: area under the curve.

[a]Cohen effect size: 0.2, small; 0.5, medium; 0.8, high.

[c]Student *t* test.

The temperature at which the first antipyretic was ingested was defined as the baseline temperature, and the mean difference in temperature for each time over 24 hours was compared between antipyretic administration patterns (Figure 3). As shown in Figure 3A, fever reduction was higher in the single group before 6 hours, and there was little difference in delta fever between the two groups after 6 hours. Comparing the antipyretic agents in the single group, IBU showed the highest fever reduction before 6 hours. However, there was no significant overall difference between agents (Figure 3B). In the combination group, the IBU-DEX combination showed the largest fever difference before 4 hours, but the ACE-IBU combination showed the largest fever difference after 4 hours. After 8 hours, the fever difference between the single and combination groups or antipyretic agents became very small.

Figure 3. Average temperature changes according to the antipyretic administration pattern for 24 hours in cases where the temperature at the time of antipyretic administration was >38 ˚C: (A) comparison of the single and combination groups and (B) comparison of the type of antipyretic agent administered.



Discussion

Principal Findings

This is the first and largest cohort study concerning the antipyretic use pattern and antipyretic efficacy between single and combination antipyretics in children using large-scale patient-generated health data (PGHD) on fever, antipyretic agents, and temperature. Most patients used single antipyretics, the most common of which was single ACE. Patients were more likely to use a combination of antipyretics at higher temperatures, and the ACE-DEX combination was the most used. However, fever reduction within 6 hours was significantly better in the single group than in the combination group, as reflected by the AUC analysis. Therefore, a combination of different antipyretics may not have a significant advantage over a single antipyretic. An additional contribution of this study is that we analyzed large-scale real-world PGHD rather than relying on data from a small, clinical-based group; this study used data derived from anyone using a mobile phone and not just from a specific hospital or location, which allowed us to analyze individual data covering a larger area.

Antipyretics in Use

For fever, ACE is the most prescribed drug, followed by IBU [23]. Both antipyretic agents inhibit prostaglandin synthetase in the hypothalamus, and both are proven to reduce fever [24]. DEX is an NSAID and a pharmacologically effective enantiomer of racemic IBU [25]. DEX can exert identical pharmacological efficacy with smaller doses instead of racemic IBU, which could potentially decrease side effects [25]. In several trials, IBU and DEX showed comparable antipyretic effects and tolerability [25,26]. ACE and IBU or DEX have different modes of action and therefore have been used in combination.

The Fever Coach app is useful to observe antipyretic use patterns in the general pediatric population. In this study, 73.5% of cases used single antipyretics, and most of them used ACE (29.5%). DEX was the second most used antipyretic at 27.0%. ACE was used in younger children and was not frequently administered at higher temperatures. This use pattern might suggest that parents thought ACE was safer but less effective in fever reduction than IBU or DEX. These results also revealed parents' perceptions of the temperature that required antipyretics. Most cases were administered antipyretics when temperatures were above 38.0 ˚C; however, 5.5% used antipyretics under 38 ˚C.

Previous studies on parents' perceptions of fever and antipyretics revealed that many parents were unfamiliar with fever standards [27-29]. In a survey of 105 parents from two emergency rooms in the United States, 81% of parents defined fever as a temperature <38 ˚C [27]. Additionally, 89% reported administering antipyretics to their children even though they seemed comfortable. A survey of 1032 children who visited tertiary hospitals in Turkey revealed that one-third of them had temperatures <37.8 ˚C [29]. These and our results suggest that parents should be educated that temperature does not determine the severity of the disease [1] and antipyretics should not be used to reduce fever but rather to relieve discomfort or pain.

Patterns of Alternating or Combination Antipyretics

In Korea, surveys on the use or combination of antipyretics are limited [28,30]. Most primary pediatricians recommend alternatively combined rather than simultaneously combined regimens due to concerns regarding additive adverse effects [1,31,32]. Although 4 hours was the most frequent interval in the previous reports, parents also reported alternating therapy every 2 hours, 3 hours, 4 hours, and 6 hours, suggesting that there was no consensus on dosing instructions.

Most cases in the combination group (95%) in our study used an ACE-based combination. ACE-based antipyretics were used in younger children more than other combinations, with an average age of 27.7 months, 28.5 months, and 29.0 months in the ACE-IBU, ACE-DEX, and ACE-IBU-DEX groups, respectively, while the average age in the IBU-DEX group was 30.0 months. Because the average age of ACE cases in the single group was younger than IBU or DEX, the combination pattern in this study might indicate that most parents used ACE first then IBU or DEX in some cases. In this case, patients were more likely to alternatively combine antipyretics at higher temperatures. The ACE-DEX combination was the most administered (12.3% of the combination group) when the temperature was between 38 and 39 ˚C, whereas 20.2% administered ACE-DEX when the temperature was >39 ˚C. This might be a use pattern reflective of the caregiver's concerns about fever and the perception that antipyretic combination therapy might be better than monotherapy to reduce fever.

Effects of Alternating or Combination Antipyretics

Using meta-analysis, IBU was equal or more efficacious in reducing fever than ACE in both pediatric and adult patients and showed similar adverse events [33]. However, previous studies have not uniformly favored combination antipyretics. An Indian study included 89 children using IBU alone or in combination with ACE in which ACE-IBU was more effective than ACE alone, but the effect was <0.5 ˚C [34]. A British study randomized 123 children who received ACE or IBU or both. They reported a temperature difference >1 ˚C between all treatments, but only a 0.35 ˚C difference was found between ACE-IBU and ACE, and 0.25 ˚C between ACE-IBU and IBU [35]. There are currently no clinical guidelines that routinely recommend the combined use of antipyretics and no consistent evidence that combination therapy results in overall improvement in clinical outcomes.

Here, the magnitude of fever reduction within 6 hours, 8 hours, 10 hours, and 12 hours after administration of the first antipyretic agent was significantly better in the single than in the combination group. Patients were more likely to use combination antipyretics at higher temperatures. However, no significant benefit was obtained by adding different antipyretic agents. This trend was consistently observed, regardless of the type of antipyretics used.

NSAIDs have a ceiling effect as analgesics in that there exists a dose beyond which there is no additional effect [36]. Higher doses do not provide additional pain relief but may increase the likelihood of side effects. ACE also demonstrates a ceiling effect in pain relief [37]. However, there is no known ceiling effect

in anti-inflammatory doses of NSAIDs or ACE [37]. Thus, the effect of combination antipyretics on fever course remains inconclusive. Based on our results, it is possible that combination antipyretics reduce fever more slowly, indicating more fever exposure during the first several hours administration of the antipyretic agent.

Adverse Effects of Alternating or Combination Antipyretics

Safety issues are concerning considering the antipyretic combination treatment. In randomized trials in adults with knee pain, drug-related adverse effects were significantly higher in subjects taking combination drugs compared with those taking IBU monotherapy (51% and 42%; *P*=.04) [38]. Moreover, the incidence of diarrhea and hepatotoxicity was significantly higher in the combination therapy group. Additionally, occasional side effects including renal failure [39] and hypothermia were observed [40]. ACE and IBU may cause acute renal failure synergistically as a result of ACE oxidative metabolites accumulating in the renal medulla during renal ischemia, which can be caused by NSAIDs [41,42].

Limitations

Regarding the limitations of this study, the dataset did not contain information related to adverse effects. In addition, it was difficult to objectively determine the frequency of hypothermia because there were many cases where the body temperature was not entered in the app after the body temperature became normal. However, the concomitant medications and causes of fever would vary between each patient. Thus, it was difficult to assess only antipyretic-related side effects. Although the frequency was very low, there is a risk of serious side effects.

Here, fever was defined as a temperature >38 ˚C, regardless of the child's age. Although fever was originally defined using rectal temperatures, this is not a readily applicable method and infrared tympanic and axillary thermometers are more commonly used [43,44]. However, this dataset did not contain information on the temperature measurement method or site, and it was difficult to analyze all relevant factors, including age and biological factors. Second, a single fever case defined in this study might not be a fever event under the same conditions. Concomitant medication or other general conditions might affect the antipyretic process. Third, since the study was based on patient-generated data, it was dependent on the user entering the data correctly and consistently, and we were unable to collect information on adverse effects. As consistent data could not be collected over time intervals, we used linear interpolation to obtain the body temperature hourly. Moreover, assuming that the change in body temperature over time is linear may have biased the results. Last, there is a possibility of selection bias depending on who used this app first. This app might have been used by caregivers who were either more sensitive or more concerned about fever. Moreover, the combination antipyretics might be more actively selected because of a previous history of fever or febrile convulsions in certain patients. Furthermore, parents with more severely ill children might have a tendency to use combined antipyretics. These limitations could be partially overcome through wearing thermometers and live monitoring equipment in a future study.

Conclusion

Antipyretic combination patterns were analyzed using real-time PGHD in a large cohort. Approximately 75% of febrile cases used single antipyretics (mostly ACE), but combination use became more frequent as the temperature increased. Single antipyretics showed faster defervescence and reduced the total exposure to fever by duration and temperature. Multiple combined antipyretic administrations did not show definite advantages over single antipyretic use in reducing fever, regardless of the different antipyretics. Moreover, there were also inappropriate uses, such as administering antipyretics at low temperatures or in triple combinations.

Thus, these data suggest that implementation of educational programs and guidelines regarding the proper management of a febrile child are needed, and a mobile app could be a useful platform for this purpose. Single antipyretics are effective in relieving discomfort in febrile children. Combination antipyretics may place children at an increased risk without additional benefit. When educating caregivers, health care providers should minimize fever phobia and emphasize the importance of monitoring the signs and symptoms of a child and improving the child's comfort in addition to the appropriate dosing of antipyretics.

Acknowledgments

We would like to thank Mobile Doctor for providing the log data of the mobile Fever Coach app.

Authors' Contributions

YRP and HK conceptualized and designed the study, interpreted the initial analysis, drafted the initial manuscript, and reviewed and revised the manuscript. JAP processed the data, performed the initial analysis, and drafted the initial manuscript. SHA and SC acquired the data and revised the manuscript critically for important intellectual content. JWS acquired the data, supervised the study, and revised the manuscript critically for important intellectual content. MK acquired and processed the data and revised the manuscript critically for important intellectual content. JHL conceptualized, designed, and supervised the study and reviewed and revised the manuscript critically for important intellectual content. HK and JHL analyzed the efficacy of antipyretic effects according to the antipyretic administration pattern. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

SC, JS, and MK are employees of Mobile Doctor Co, Ltd. All other authors declare no conflicts of interest.

Multimedia Appendix 1
STROBE Statement for manuscript.
[DOCX File , 33 KB - mhealth_v9i5e21668_app1.docx]

Multimedia Appendix 2
Descriptive statistics and children's temperature records according to the antipyretic drug administration pattern and antipyretic ingredients.
[DOCX File , 21 KB - mhealth_v9i5e21668_app2.docx]

Multimedia Appendix 3
Comparison of the area under the curve between the administration pattern according to antipyretic ingredient.
[DOCX File , 19 KB - mhealth_v9i5e21668_app3.docx]

References

1. Kim JS. Childhood fever management: current practice vs evidence. Child Health Nurs Res 2016 Apr 30;22(2):126-136. [doi: 10.4094/chnr.2016.22.2.126]
2. Kwak BG, Jang HO. Clinical analysis of febrile infants and children presenting to the pediatric emergency department. Korean J Pediatr 2006;49(8):839-844. [doi: 10.3345/kjp.2006.49.8.839]
3. Section on Clinical Pharmacology and Therapeutics, Sullivan JE, Farrar HC. Fever and antipyretic use in children. Pediatrics 2011 Mar;127(3):580-587. [doi: 10.1542/peds.2010-3852] [Medline: 21357332]
4. Mayoral CE, Marino RV, Rosenfeld W, Greensher J. Alternating antipyretics: is this an alternative? Pediatrics 2000 May;105(5):1009-1012. [doi: 10.1542/peds.105.5.1009] [Medline: 10790455]
5. Paul IM, Sturgis SA, Yang C, Engle L, Watts H, Berlin CM. Efficacy of standard doses of Ibuprofen alone, alternating, and combined with acetaminophen for the treatment of febrile children. Clin Ther 2010 Dec;32(14):2433-2440 [FREE Full text] [doi: 10.1016/j.clinthera.2011.01.006] [Medline: 21353111]
6. Manthripragada AD, Zhou EH, Budnitz DS, Lovegrove MC, Willy ME. Characterization of acetaminophen overdose-related emergency department visits and hospitalizations in the United States. Pharmacoepidemiol Drug Saf 2011 Aug;20(8):819-826. [doi: 10.1002/pds.2090] [Medline: 21294217]
7. Nagai J, Uesawa Y, Shimamura R, Kagaya H. Characterization of the adverse effects induced by acetaminophen and nonsteroidal anti-inflammatory drugs based on the analysis of the Japanese adverse drug event report database. Clin J Pain 2017 Aug;33(8):667-675 [FREE Full text] [doi: 10.1097/AJP.0000000000000457] [Medline: 27898459]
8. Arıkan Z, Tekşam Ö, Kara A, Kale G. Determining causes and frequency of misdosing of antipyretics in patients presenting with fever to pediatric emergency. Turkish Arch Pediatr 2012;47(2):114-118. [doi: 10.4274/tpa.716]
9. Li SF, Lacher B, Crain EF. Acetaminophen and ibuprofen dosing by parents. Pediatr Emerg Care 2000 Dec;16(6):394-397. [doi: 10.1097/00006565-200012000-00003] [Medline: 11138879]
10. Rodieux F, Piguet V, Desmeules J, Samer CF. Safety issues of pharmacological acute pain treatment in children. Clin Pharmacol Ther 2019 May;105(5):1130-1138. [doi: 10.1002/cpt.1358] [Medline: 30648741]
11. Wu A. Minimizing medication errors in pediatric patients. US Pharm. URL: https://www.uspharmacist.com/article/minimizing-medication-errors-in-pediatric-patients [accessed 2021-04-23]
12. Blieden M, Paramore LC, Shah D, Ben-Joseph R. A perspective on the epidemiology of acetaminophen exposure and toxicity in the United States. Expert Rev Clin Pharmacol 2014 May;7(3):341-348. [doi: 10.1586/17512433.2014.904744] [Medline: 24678654]
13. Nabulsi MM, Tamim H, Mahfoud Z, Itani M, Sabra R, Chamseddine F, et al. Alternating ibuprofen and acetaminophen in the treatment of febrile children: a pilot study [ISRCTN30487061]. BMC Med 2006 Mar 04;4:4 [FREE Full text] [doi: 10.1186/1741-7015-4-4] [Medline: 16515705]
14. Kramer LC, Richards PA, Thompson AM, Harper DP, Fairchok MP. Alternating antipyretics: antipyretic efficacy of acetaminophen versus acetaminophen alternated with ibuprofen in children. Clin Pediatr (Phila) 2008 Nov;47(9):907-911. [doi: 10.1177/0009922808319967] [Medline: 18539869]
15. Hay AD, Costelloe C, Redmond NM, Montgomery AA, Fletcher M, Hollinghurst S, et al. Paracetamol plus ibuprofen for the treatment of fever in children (PITCH): randomised controlled trial. BMJ 2008 Sep 02;337:a1302 [FREE Full text] [doi: 10.1136/bmj.a1302] [Medline: 18765450]
16. Craig JV, Lancaster GA, Taylor S, Williamson PR, Smyth RL. Infrared ear thermometry compared with rectal thermometry in children: a systematic review. Lancet 2002 Aug 24;360(9333):603-609. [doi: 10.1016/S0140-6736(02)09783-0] [Medline: 12241932]

17. Davis T. NICE guideline: feverish illness in children—assessment and initial management in children younger than 5 years. Arch Dis Child Educ Pract Ed 2013 Dec;98(6):232-235. [doi: 10.1136/archdischild-2013-304792] [Medline: 24046395]
18. Low JG, Sung C, Wijaya L, Wei Y, Rathore APS, Watanabe S, et al. Efficacy and safety of celgosivir in patients with dengue fever (CELADEN): a phase 1b, randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet Infect Dis 2014 Aug;14(8):706-715. [doi: 10.1016/S1473-3099(14)70730-3] [Medline: 24877997]
19. Morris PE, Promes JT, Guntupalli KK, Wright PE, Arons MM. A multi-center, randomized, double-blind, parallel, placebo-controlled trial to evaluate the efficacy, safety, and pharmacokinetics of intravenous ibuprofen for the treatment of fever in critically ill and non-critically ill adults. Crit Care 2010;14(3):R125 [FREE Full text] [doi: 10.1186/cc9089] [Medline: 20591173]
20. Mullins ME, Empey M, Jaramillo D, Sosa S, Human T, Diringer MN. A prospective randomized study to evaluate the antipyretic effect of the combination of acetaminophen and ibuprofen in neurological ICU patients. Neurocrit Care 2011 Dec;15(3):375-378. [doi: 10.1007/s12028-011-9533-8] [Medline: 21503807]
21. Kim M, Yune S, Chang S, Jung Y, Sa SO, Han HW. The Fever Coach mobile app for participatory influenza surveillance in children: usability study. JMIR Mhealth Uhealth 2019 Oct 17;7(10):e14276 [FREE Full text] [doi: 10.2196/14276] [Medline: 31625946]
22. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. New York: Taylor & Francis Group; 2013.
23. Kanabar DJ. A clinical and safety review of paracetamol and ibuprofen in children. Inflammopharmacology 2017 Feb;25(1):1-9 [FREE Full text] [doi: 10.1007/s10787-016-0302-3] [Medline: 28063133]
24. Hersh EV, Moore PA, Ross GL. Over-the-counter analgesics and antipyretics: a critical assessment. Clin Ther 2000 May;22(5):500-548. [doi: 10.1016/S0149-2918(00)80043-0] [Medline: 10868553]
25. Yoon JS, Jeong D, Oh J, Lee KY, Lee HS, Koh YY, et al. The effects and safety of dexibuprofen compared with ibuprofen in febrile children caused by upper respiratory tract infection. Br J Clin Pharmacol 2008 Dec;66(6):854-860 [FREE Full text] [doi: 10.1111/j.1365-2125.2008.03271.x] [Medline: 19032727]
26. Rahlfs VW, Stat C. Reevaluation of some double-blind, randomized studies of dexibuprofen (Seractil): a state-of-the-art overview. Studies in patients with lumbar vertebral column syndrome, rheumatoid arthritis, distortion of the ankle joint, gonarthrosis, ankylosing spondylitis, and activated coxarthrosis. J Clin Pharmacol 1996 Dec;36(12 Suppl):33S-40S. [Medline: 9013382]
27. Wallenstein MB, Schroeder AR, Hole MK, Ryan C, Fijalkowski N, Alvarez E, et al. Fever literacy and fever phobia. Clin Pediatr (Phila) 2013 Mar;52(3):254-259. [doi: 10.1177/0009922812472252] [Medline: 23349363]
28. Choi A, Kim JS. Fever phobia: a survey of children's parents in a pediatric outpatient clinic. Child Health Nurs Res 2014;20(2):113. [doi: 10.4094/chnr.2014.20.2.113]
29. Polat M, Kara S, Tezer H, Tapısız A, Derinöz O, Dolgun A. A current analysis of caregivers' approaches to fever and antipyretic usage. J Infect Dev Ctries 2014 Mar 13;8(3):365-371 [FREE Full text] [doi: 10.3855/jidc.3904] [Medline: 24619269]
30. Kwak YH, Kim DK, Jang HY, Kim JJ, Ryu J, Oh SB, et al. Fever phobia in Korean caregivers and its clinical implications. J Korean Med Sci 2013 Nov;28(11):1639-1644 [FREE Full text] [doi: 10.3346/jkms.2013.28.11.1639] [Medline: 24265528]
31. Demir F, Sekreter O. Knowledge, attitudes and misconceptions of primary care physicians regarding fever in children: a cross sectional study. Ital J Pediatr 2012 Sep 05;38:40 [FREE Full text] [doi: 10.1186/1824-7288-38-40] [Medline: 22950655]
32. Wright AD, Liebelt EL. Alternating antipyretics for fever reduction in children: an unfounded practice passed down to parents from pediatricians. Clin Pediatr (Phila) 2007 Mar;46(2):146-150. [doi: 10.1177/0009922806293922] [Medline: 17325087]
33. Pierce CA, Voss B. Efficacy and safety of ibuprofen and acetaminophen in children and adults: a meta-analysis and qualitative review. Ann Pharmacother 2010 Mar;44(3):489-506. [doi: 10.1345/aph.1M332] [Medline: 20150507]
34. Lal A, Gomber S, Talukdar B. Antipyretic effects of nimesulide, paracetamol and ibuprofen-paracetamol. Indian J Pediatr 2000 Dec;67(12):865-870. [doi: 10.1007/BF02723945] [Medline: 11262983]
35. Erlewyn-Lajeunesse MDS, Coppens K, Hunt LP, Chinnick PJ, Davies P, Higginson IM, et al. Randomised controlled trial of combined paracetamol and ibuprofen for fever. Arch Dis Child 2006 May;91(5):414-416 [FREE Full text] [doi: 10.1136/adc.2005.087874] [Medline: 16464962]
36. Becker DE. Pain management: part 1: managing acute and postoperative dental pain. Anesth Prog 2010;57(2):67-78 [FREE Full text] [doi: 10.2344/0003-3006-57.2.67] [Medline: 20553137]
37. Ong CKS, Lirk P, Tan CH, Seymour RA. An evidence-based update on nonsteroidal anti-inflammatory drugs. Clin Med Res 2007 Mar;5(1):19-34 [FREE Full text] [doi: 10.3121/cmr.2007.698] [Medline: 17456832]
38. Doherty M, Hawkey C, Goulder M, Gibb I, Hill N, Aspley S, et al. A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. Ann Rheum Dis 2011 Sep;70(9):1534-1541. [doi: 10.1136/ard.2011.154047] [Medline: 21804100]
39. Moghal NE, Hegde S, Eastham KM. Ibuprofen and acute renal failure in a toddler. Arch Dis Child 2004 Mar;89(3):276-277 [FREE Full text] [doi: 10.1136/adc.2002.024141] [Medline: 14977711]
40. Richardson J, Sills J. Hypothermia following fever. Arch Dis Child 2004 Dec;89(12):1177 [FREE Full text] [doi: 10.1136/adc.2004.055376] [Medline: 15557062]

41. McIntire SC, Rubenstein RC, Gartner JC, Gilboa N, Ellis D. Acute flank pain and reversible renal dysfunction associated with nonsteroidal anti-inflammatory drug use. Pediatrics 1993 Sep;92(3):459-460. [Medline: 8361805]
42. Eguia L, Materson BJ. Acetaminophen-related acute renal failure without fulminant liver failure. Pharmacotherapy 1997;17(2):363-370. [Medline: 9085330]
43. Craig JV, Lancaster GA, Williamson PR, Smyth RL. Temperature measured at the axilla compared with rectum in children and young people: systematic review. BMJ 2000 Apr 29;320(7243):1174-1178 [FREE Full text] [Medline: 10784539]
44. Berksoy E, Bıcılıoğlu Y, Gökalp G, Bal A. Comparison of infrared tympanic, non-contact infrared skin, and axillary thermometer to rectal temperature measurements in a pediatric emergency observation unit. Int J Clin Exp Med 2018;11(2):567-573.

Abbreviations

ACE: acetaminophen
AUC: area under the curve
DEX: dexibuprofen
IBU: ibuprofen
NSAID: nonsteroidal anti-inflammatory drug
PGHD: patient-generated health data

Edited by G Eysenbach; submitted 22.06.20; peer-reviewed by S Musy, CM Choi, J Buekers; comments to author 28.10.20; revised version received 27.01.21; accepted 11.04.21; published 26.05.21.

Please cite as:
Park YR, Kim H, Park JA, Ahn SH, Chang S, Shin JW, Kim M, Lee JH
Comparative Analysis of Single and Combined Antipyretics Using Patient-Generated Health Data: Retrospective Observational Study
JMIR Mhealth Uhealth 2021;9(5):e21668
URL: https://mhealth.jmir.org/2021/5/e21668
doi:10.2196/21668
PMID:34037528

Original Paper

Meal-time Smartphone Use in an Obesogenic Environment: Two Longitudinal Observational Studies

Joceline Y Y Yong[1], BA; Eddie M W Tong[2], PhD; Jean C J Liu[1,3], PhD

[1]Division of Social Sciences, Yale–National University of Singapore College, Singapore, Singapore

[2]Department of Psychology, National University of Singapore, Singapore, Singapore

[3]Neuroscience and Behavioral Disorders Programme, Duke-National University of Singapore Medical School, Singapore, Singapore

Corresponding Author:
Jean C J Liu, PhD
Division of Social Sciences
Yale–National University of Singapore College
16 College Avenue West #01-220
Singapore, 128527
Singapore
Phone: 65 6601 3694
Email: jeanliu@yale-nus.edu.sg

Abstract

Background: Despite a large volume of research on the impact of other digital screens (eg, televisions) on eating behavior, little is known about the nature and impact of mealtime smartphone use.

Objective: We investigated how smartphones are used in everyday meals, whether phone users differ according to mealtime phone use patterns, and whether specific phone functions (particularly food photography) would affect the amount and enjoyment of food eaten.

Methods: Across 2 studies, we used the experience sampling method to track 1780 meals in situ. In study 1, a total 137 young adults reported on their mealtime smartphone use 3 times per day over 7 consecutive days. This corresponded to each main meal, with participants recording whether they used their phones and what phone functions they engaged in while eating. In study 2, a total of 71 young adults were similarly tracked for 3 meals per day over 7 days. Across the week, participants' meals were randomized to 1 of 3 smartphone conditions: food photography while eating, nonfood photography while eating, or no phone use. As the outcome measures, participants reported on the amount and enjoyment of food they ate.

Results: During the week-long tracking, most participants (110/129, 85.3%) recorded at least one instance of mealtime smartphone use, with an average frequency of 1 in 3 meals where phones were used (27.1%; 95% CI 23.6-30.6). Unlike traditional digital screens, mealtime phone use encompassed a wide range of social and nonsocial activities. Further, specific forms of phone use behaviors influenced food intake in different ways. Specifically, in study 2, participants showed the typical pattern of increased food intake across the day when they engaged in nonfood photography during a meal (P<.001); however, this pattern was disrupted when they engaged in food photography (P=.73).

Conclusions: Our findings underscore the prevalence and multifaceted nature of mealtime phone use, distinguishing mobile phones from traditional forms of digital screens.

Trial Registration: ClinicalTrials.gov NCT03299075; https://www.clinicaltrials.gov/ct2/show/NCT03299075 and ClinicalTrials.gov NCT03346785; https://clinicaltrials.gov/ct2/show/NCT03346785

(JMIR Mhealth Uhealth 2021;9(5):e22929) doi:10.2196/22929

KEYWORDS

screen time; mobile phones; technology; obesogenic environment; young adults

Introduction

A large body of research underscores how using digital screens (eg, television, computers) can predispose an individual to obesity [1-3]. For example, epidemiological studies consistently report that increased screen time is associated with decreased physical activity and a higher BMI [1-6]. In laboratory studies, those who watch television or play computer games while eating

were found to consume more calories [7-9], be less aware of how much food they have eaten [9,10], and rely more heavily on external cues to determine satiety [11]. Accordingly, various regulatory and industry groups (eg, the American Academy of Pediatrics) have issued guidelines advocating for digital screens to be put aside during meal times [12-14].

Although these guidelines have been based primarily on television research, mobile phones (or smartphones) have also been linked to weight gain [15,16] and are increasingly being used at the dinner table [17-19]. This is unsurprising given that smartphone penetration rates have risen globally [20], particularly among younger individuals (aged 18-34 years), who report greater use of smartphones, the internet, and social media (relative to older adults aged over 50 years) [21]. Among youths, the rate of smartphone ownership reaches 84%-99% in countries with advanced economies [21].

Despite the ubiquity of mobile devices, there have been few studies on mealtime phone use, and the impact of this phenomenon remains unclear. On the one hand, a mobile phone resembles other digital screens in its sedentary usage [22,23] and ability to distract the phone user [24,25]. On the other hand, the vast number of smartphone functions may mean that someone using a phone while eating may engage with food in ways that differ from traditional digital screens [26].

Of note, one particular form of mealtime phone use has received heavy criticism, that of food photography, where the meal is photographed and the images are shared on social media. Although the capacity to take photographs pre-dates mobile phones, incorporating camera functions into the phone has transformed the extent and means by which we capture our meals. Correspondingly, food pictures now rank among the top categories of images uploaded on the photo-sharing platform Instagram.

Food photography is illustrative of how mealtime phone use can introduce new rituals to the dinner table. Unlike other forms of digital screens (or indeed other forms of phone use), taking photographs involves direct engagement with the meal—arranging the food, taking a photograph, applying a filter, and posting the photo on social media. In the public domain, restaurateurs have been so concerned that this would detract from the eating experience that they moved to ban it [27]. Similarly, this act was singled out in clinical circles as being potentially pathological [28]. Together, these concerns point to a broader conversation on how advances in mobile phone technology have the potential to transform eating habits.

In light of these developments, there is a need for empirical research to understand the place of phones at the dinner table and how they contribute to an obesogenic environment [29]. To this end, we conducted two 2 studies (NCT03299075 and NCT03346785) using the experience sampling method to capture mealtime phone use as it occurred in day-to-day routines (a method frequently applied in health psychology research [30-32]). By tracking usage patterns in real time, this method had high ecological validity and circumvented recollection biases associated with traditional surveys [33]. This is particularly important when studying phone use, as users have been found to provide poor self-reported estimates on how frequently and in what context they use their phones [34]. Using experience sampling, we thus investigated how mobile phones are used in everyday meals (study 1), whether the characteristics of phone users differ according to mealtime phone use patterns (study 1), and how specific phone functions, particularly food photography, may affect the amount and enjoyment of food eaten (study 2).

Methods

Recruitment

Across both studies, participants were recruited from the National University of Singapore via advertisements. All participants provided informed consent at the start of the study and were reimbursed SGD $5 (study 1) or SGD $10 (study 2) for their time. Protocols were approved by the National University of Singapore's Institutional Review Board (A-15-170) and were preregistered at ClinicalTrials.gov (NCT03299075 and NCT03346785).

Participants

In the first study, participants were 137 young adults who met the following eligibility criteria: ownership of a smartphone, aged between 18-30 years, with no history of medical or psychiatric disorders (including eating disorders), and nonsmoker status. The second study involved 71 young adults recruited using the same eligibility criteria as study 1. (For both studies, Multimedia Appendix 1 provides a detailed flow diagram of participant selection and dropout.)

Procedures

As baseline measures, participants self-reported demographic information (age, gender, ethnicity). They also reported 3 health-related metrics: their height and weight (used to derive their BMI), and their frequency of engaging in vigorous physical activity (defined as the number of times, during a single week, they engaged in activities where they "worked up a sweat").

Participants also completed the Dutch Eating Behavior Questionnaire (DEBQ) [35], a 33-item scale that determined whether participants deliberately restrained their eating (eg, by refusing food or drink because of weight concerns), ate in response to emotions (eg, when irritated), or ate in response to the external environment (eg, when food smelled and looked good). Finally, participants reported on their habitual phone use patterns [26] (ie, contexts during which participants used their phones, phone functions used, and social media activity).

For the experience sampling component, participants were contacted daily for 7 days (Monday to Sunday) via the Facebook Messenger app (Facebook, Inc) on their mobile phones. Each day, participants received 3 prompts sent at customized schedules coinciding with their regular meal times (breakfast, lunch, and dinner). These prompts were delivered at the following median times: 9:15 AM (study 1) and 8:45 AM (study 2) for breakfast, 12:45 PM (studies 1 and 2) for lunch, and 6:45 PM (studies 1 and 2) for dinner

All messages were delivered via a Python-programmed Facebook bot, and responses were recorded through the Messenger platform with a 30-minute time-out window (median

response rate for study 1: 16/21,76%; for study 2: 21/21, 100%). A full list of questions asked of participants can be found in Multimedia Appendix 2.

Experience Sampling: Questions on Mealtime Phone Use (Study 1)

As study 1 was designed to understand everyday mealtime phone use, participants were asked at each prompt whether they had eaten in the last 15 minutes. In this manner, we captured 1140 meals within the day-to-day routines of 129 free-living participants.

If participants reported that they had eaten, they then indicated whether they used their phones during the meal and what phone functions they used. As a pilot for study 2, participants also recorded their consumption patterns at each prompt (Multimedia Appendix 3).

Experience Sampling: Experiment on the Impact of Phone Use Patterns (Study 2)

In study 2, we sought to investigate the causal impact of phone use, particularly that of mealtime food photography, on eating behaviors. As food photography involves a set of rituals in naturalistic settings (eg, arranging the food, taking photos from different angles, applying photo filters) [36], we again used experience sampling to study this phenomenon in situ. To this end, we first asked participants at each prompt if they were going to eat within the next 15 minutes, allowing us to capture 640 meals from 70 participants.

To allow causal inferences, we then randomly assigned meals to 1 of 3 phone use conditions. If participants had indicated that they were about to eat, they were sent one of the following instructions: (1) take a photograph of their food as if they were doing so for the photo-sharing application Instagram (food photography condition), (2) take a photograph of their surroundings (eg, furniture, decorations) as if for Instagram (matched nonfood photography condition), (3) or refrain from phone use (no phone condition). These instructions were randomized within the week, with each participant experiencing all 3 conditions. Critically, a follow-up message was then sent 30 minutes after the instructions to verify compliance: participants were either asked to upload the photograph they took (for the food and nonfood photography conditions; see Multimedia Appendix 4 for sample images) or to indicate whether they had used their phones while eating (for the no phone condition).

Finally, as the primary outcome measures, we tracked participants' enjoyment of the meal and the amount they had eaten. At the end of each prompt, participants rated how much they enjoyed the food using a 7-point scale anchored with "1 = not at all" and "7 = very much". To assess the amount eaten, we took a relative measurement approach based on food consumption studies that had employed experience sampling [37]. Using a 7-point scale anchored with "1 = less than normal" and "7 = more than normal," participants rated how much they had eaten. Self-reported portions increased across the day (from breakfast to lunch to dinner; main effect of meal type on amount: $t_{276,66}$=3.12; P=.002), mirroring the diurnal intake rhythms that have been observed in food diary studies of free-living individuals [38-40] and thus validating this approach (see Multimedia Appendix 3 for validation analyses).

Statistical Analyses

To characterize mealtime phone use patterns (study 1), observations are summarized with counts and percentages. Characteristics of phone users themselves are summarized with means (SD) or counts and percentages, with subgroups of phone users compared using *t* tests.

For Study 2, we analyzed all instances where participants complied to the experimental instructions. Following the analysis strategy of screen-time studies, we planned a set of orthogonal contrasts to first assess the impact of phone use (vs no phone use) [9] and the impact of different forms of phone use (food vs nonfood photography) [41]. Applying contrast coding, we coded whether participants had been assigned to use their phones during the meal (phone use variable: –1=no; 0.5=yes, food photography; 0.5=yes, nonfood photography), and, if they had, we coded what form of phone photography they had engaged in (photography variable: 0=no phone; –1=nonfood photography; 1=food photography). For each outcome measure (enjoyment and amount), we ran a linear mixed-effects model with time (centered and divided by 3 to put the unit in days), phone use, photography, and meal type (breakfast, lunch, or dinner), with the interaction between each phone variable and meal type entered as fixed effects. Random intercepts accounted for correlated data due to repeated measures, and the type 1 decisionwise error rate was controlled at α=.05.

Across both studies, analyses were conducted using SPSS 25 (IBM Corp) and R 3.4.0 (R Foundation for Statistical Computing).

Results

Baseline Participant Characteristics

In study 1, the participants had a mean age of 21.68 years (SD 2.07), a mean BMI of 20.89 (SD 2.80), and were predominantly of Asian ethnicity: 85.3%(110/129) self-identified as Chinese, 3.8% (5/129) Indian, 3.1% (4/129) Malay, 1.6% (2/129) Eurasian, and 6.2% (8/129) as another ethnicity. Moreover, 73.6% (95/129) identified as female. In study 2, the participants had a mean age of 22.29 years (SD 2.36), a mean BMI of 21.80 (SD 3.22), and were predominantly of Asian ethnicity: 89% (62/70) self-identified as Chinese, 4% (3/70) Indian, 1% (1/70) Malay, and 6% (4/70) another ethnicity. Moreover, 57 of the 70 participants (81%) identified as female.

Across both studies, the average participant's BMI fell within the normal range [42]. Relative to the resident population, both samples had a higher proportion of females and persons of Chinese ethnicity (>10% difference) [43].

How Mobile Phones Are Used in Everyday Meals (Study 1)

In the baseline phone-use questionnaire, participants self-reported that they were "likely" to use their phones during a meal (mean rating 3.74/5; 95% CI 3.58-3.91; see Figure 1).

This rating was quantified by the experience sampling data: during the week of in-depth monitoring, the vast majority of participants (110/129, 85.3%) recorded at least one instance of mealtime phone use. On average, participants used their phones during 1 of 3 meals (27.1%; 95% CI 23.6-30.6).

In terms of how phones were used, our week-long monitoring captured a wide range of social (eg, talking on the phone) and nonsocial phone activities (eg, listening to music) that participants engaged in while eating. These patterns broadly mapped onto those outside the eating context (as reported in previous survey studies [26,44,45] and by participants in the baseline questionnaires; Figure 2). Namely, across all contexts, a similar variety and rank ordering of phone functions were used (Kendall τ=0.64; P=.03).

However, mealtime phone patterns were distinct in one important aspect. Although messaging or social networking was the top-ranked functions across all contexts, participants were far more likely to use this feature while eating than to use any other phone function. Indeed, 86.5% (313/362; 95% CI 83.5-90.4) of all mealtime phone use episodes captured involved messaging or social networking, which was 8 times the frequency of the next most widely used function, browsing websites (39/362, 11%; 95% CI 8.7-15.3). Correspondingly, mealtime phone use was more likely to be social than nonsocial in nature. Outside the eating context, however, participants self-reported that they used a range of social and nonsocial phone functions comparably (mean rating of 4.0-4.88 for the likelihood of listening to music, taking photos or videos, watching videos, browsing websites, and messaging or social networking).

Figure 1. In a baseline questionnaire on habitual phone use, participants reported how likely they were to use their phones in each context. A higher score corresponds to greater likelihood, and horizontal lines represent the 95% CI for the mean. When participants were then monitored for 1 week, mealtime phone use was observed in approximately 9 of 10 participants (in an average of 1 in 3 meals).

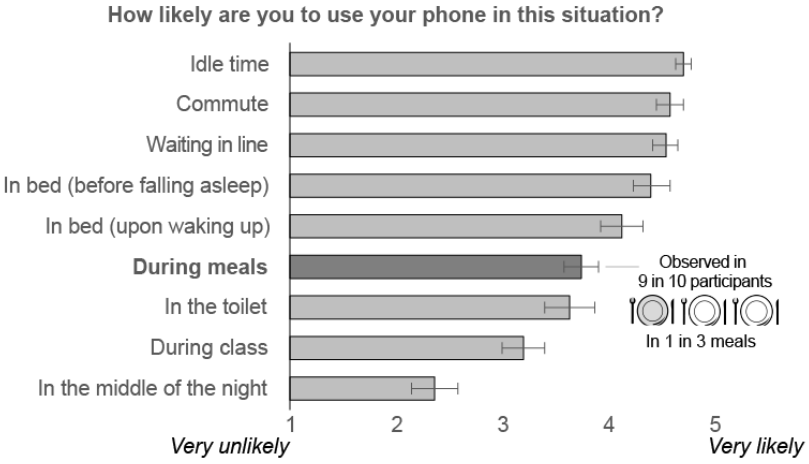
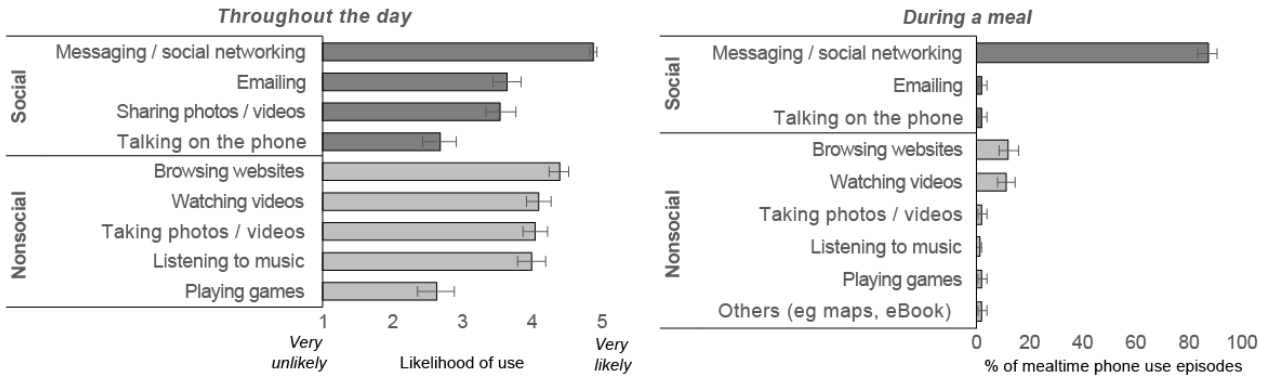


Figure 2. (Left) In a baseline questionnaire on habitual phone use, participants reported how likely they were to use each phone function on a regular day. A higher score corresponds to greater likelihood, and horizontal lines represent the 95% CI for the mean. (Right) Participants were then monitored closely for 1 week; the graph on the right depicts the percent of mealtime phone use episodes where each phone activity was recorded. Horizontal lines represent the 95% CI for each percentage.



Characteristics of Phone Users as a Function of Mealtime Usage (Study 1)

The experience sampling method captured large individual differences in mealtime phone use patterns. Across the week of observation, we recorded the full range of 0%-100% of meals where phones were used (Figure 3; Multimedia Appendix 5 shows how this was not merely an artifact of the number of meals captured per participant). Accordingly, we conducted exploratory subgroup analyses to examine whether “chronic”

mealtime phone users (the top 15% of participants in phone usage, corresponding to 50%-100% of meals with phone use) differed in characteristics from "regular" mealtime phone users (the remaining 85% of participants, corresponding to <50% of meals with phone use).

As shown in Table 1, chronic mealtime phone users tended to be older (t_{127}=–1.94; *P*=.05) and to use their phones more frequently for watching videos (t_{127}=–2.28; *P*=.02). However, we found no evidence that chronic users differed in BMI, levels of vigorous physical activity each week, or habitual eating patterns as measured by the DEBQ (smallest *P*=.19).

Figure 3. Box-plot depicting the distribution of mealtime phone use frequency captured across 1 week of naturalistic monitoring. The bottom, midline, and top of the box represent the 25th, 50th, and 75th percentiles, respectively, and chronic users are represented in the shaded gray area (top 15% of participants, corresponding to ≥50% of meals with phone use).

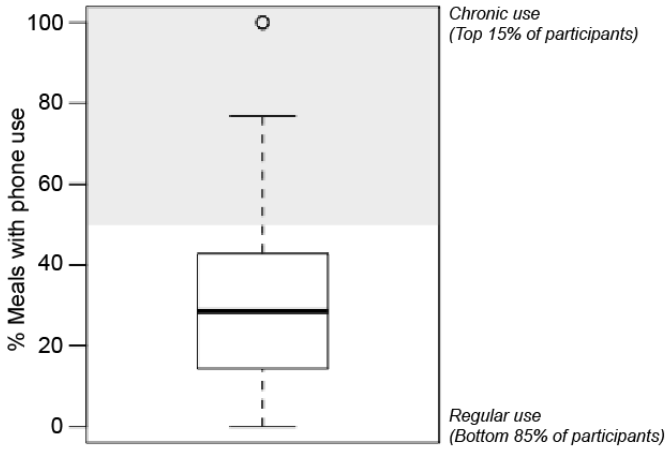


Table 1. Participant characteristics as a function of mealtime phone use patterns[a].

Characteristic	Chronic mealtime phone users (n=22)	Regular mealtime phone users (n=107)	Test statistic (*P* value)
Observed mealtime phone use			
Proportion of meals with phone use	59.83 (11.77)	20.31 (13.73)	–12.58 (*<.001*)[b]
Demographics[c]			
Age (years)	22.45 (2.44)	21.52 (1.96)	–1.94 (.05)
Gender, n (%)			0.18 (.67)[d]
Female	17 (77.27)	78 (72.90)	
Male	5 (22.73)	29 (27.10)	
Ethnicity, n (%)			2.51 (.64)[d]
Chinese	19 (86.36)	91 (85.05)	
Malay	0 (0.00)	4 (3.74)	
Indian	1 (4.55)	4 (3.74)	
Eurasian	1 (4.55)	1 (0.93)	
Others	1 (4.55)	7 (6.54)	
Phone brand, n (%)			2.63 (.62)[d]
Apple	15 (68.18)	74 (69.16)	
Samsung	4 (18.18)	19 (17.76)	
Others	3 (13.64)	14 (13.08)	
Health-related variables[c]			
BMI	20.18 (2.27)	21.04 (2.88)	1.31 (.19)
Frequency of vigorous physical activity, n (%)			0.30 (.86)[d]
0-2 times/week	15 (0.68)	68 (63.55)	
3-4 times/week	6 (27.27)	31 (28.97)	
5-7 times/week	1 (4.55)	8 (7.48)	
Dutch Eating Behavior Questionnaire[c]			
Restrained eating score	2.30 (0.66)	2.48 (0.83)	0.95 (.34)
Emotional eating score	2.50 (0.96)	2.61 (0.80)	0.56 (.58)
External eating score	3.39 (0.62)	3.32 (0.54)	0.57 (.57)
Self-reported phone use habits[c]			
Social phone functions engaged in			
Messaging or social networking	4.86 (0.35)	4.89 (0.32)	0.32 (.75)
Emailing	3.82 (1.05)	3.60 (1.12)	–0.85 (.40)
Sharing photos or videos	3.50 (1.34)	3.56 (1.15)	–0.19 (.85)
Talking on the phone	2.73 (1.32)	2.65 (1.30)	–0.24 (.81)
Nonsocial phone functions engaged in			
Browsing websites	4.50 (0.60)	4.36 (0.79)	–0.76 (.45)
Watching videos	4.55 (0.51)	4.01 (1.08)	2.28 (*.02*)
Taking photos / videos	3.91 (1.07)	4.07 (0.96)	0.68 (.50)
Listening to music	3.82 (1.14)	4.04 (1.13)	0.83 (.41)
Playing games	2.82 (1.76)	2.58 (1.43)	0.80 (.43)

Characteristic	Chronic mealtime phone users (n=22)	Regular mealtime phone users (n=107)	Test statistic (*P* value)
Typical context for phone use			
Idle time	4.82 (0.40)	4.67 (0.47)	–1.35 (.18)
During commute	4.73 (0.55)	4.55 (0.76)	–1.04 (.30)
While waiting in line	4.64 (0.49)	4.50 (0.69)	–0.85 (.40)
In bed before falling asleep	4.64 (0.73)	4.35 (0.98)	–1.31 (.20)
In bed when waking up	4.23 (1.02)	4.09 (1.15)	–0.51 (.62)
During meals	4.09 (0.61)	3.67 (1.00)	–1.89 (.06)
In the toilet	3.45 (1.57)	3.66 (1.30)	0.66 (.51)
During class time	3.32 (1.29)	3.18 (1.11)	–0.53 (.60)
In the middle of the night	2.55 (1.54)	2.31 (1.22)	–0.79 (.43)
Social networking involvement			
Number of Instagram followers	413.0 (243.81)	534.3 (735.23)	0.71 (.48)
Number of Instagram accounts followed	598.3 (471.74)	487.1 (261.84)	–1.45 (.15)

[a]Unless otherwise stated, the data are reported as means (SD), and the test statistic refers to the *t* statistic.

[b]Italics indicate *P* value <.05.

[c]Based on responses to the baseline questionnaires.

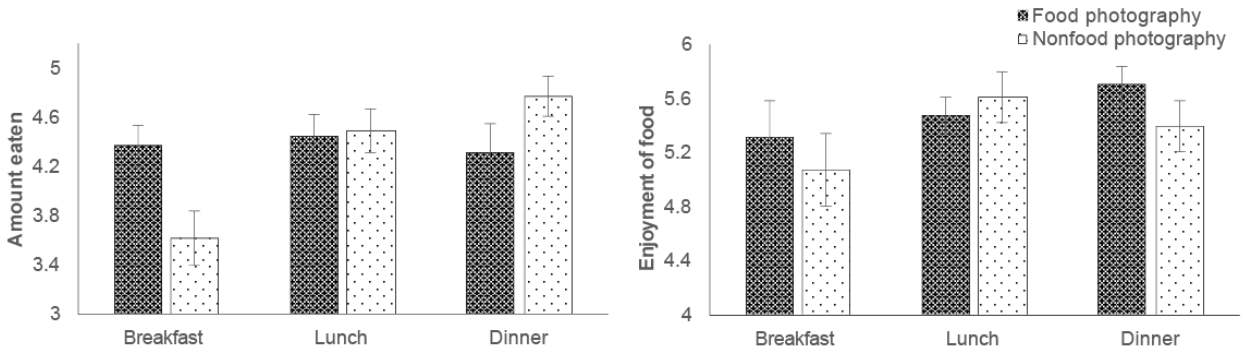
[d]Chi-square statistic reported.

Impact of Mealtime Phone Use Patterns on Eating Behaviors (Study 2)

In Study 2, we used linear mixed-effects models to examine the causal impact of phone use patterns, particularly mealtime food photography, on the amount and enjoyment of food eaten. For food intake, we observed a significant interaction between photography condition and meal type ($t_{324,60}$; =–2.61; *P*=.009). As shown in Figure 4, when participants photographed nonfood items during a meal, they showed the typical increase of food intake from breakfast to lunch and dinner (effect of meal for nonfood condition: $t_{93,81}$=4.15; *P*<.001). On the other hand, when they took mealtime photographs of their food, there was no significant effect of the type of meal consumed (effect of meal for food condition: $t_{106,34}$=0.35; *P*=.73). In other words, when participants took photographs of their food, we found no evidence that they discriminated between breakfast, lunch, or dinner in the amount of food eaten. Finally, there was no significant main or interaction effect involving phone use more generally (that is, whether or not participants used their phones during the meal; smallest *P*=.25).

In terms of the enjoyment of food eaten, there were no significant effects involving photography condition (smallest *P*=.72) or general phone use (smallest *P*=.48; Figure 4; for the 2 mixed-effects models, detailed model output including parameter estimates can be found in Multimedia Appendix 6.)

Figure 4. Mean amount and enjoyment of food eaten at each meal (breakfast, lunch, or dinner), plotted as a function of whether participants engaged in food photography or nonfood photography. A higher score corresponds to greater enjoyment or amount eaten, and vertical lines represent 1 SE of the mean.



Discussion

Principal Findings

In this series of studies, we documented how mobile phones have found a place at the modern-day dining table. During 1 week of in-depth monitoring, nearly 9 in 10 of our participants recorded at least one instance of mealtime phone use. This widespread prevalence points to new norms in eating habits, giving impetus for research on the phenomenon.

As the first step in this direction, we recorded what activities participants engaged in when they used their phones during a meal. As compared to traditional digital screens where participants perform only one function (eg, watching television), mealtime phone use entails a wide range of activities that are both social (eg, sending messages to peers) and nonsocial (eg, watching videos) in nature. This has implications for clinical guidelines and research, both of which have classified phone use as "screen time" without differentiating mobile phones from other forms of digital screens, or between various mobile phone activities [12-14]. By contrast, our findings suggest a need to reframe mealtime phone use. Given the wide range of activities engaged in, phone use may be more profitably viewed as multidimensional rather than monolithic, with consideration given to the specific function a phone user engages in.

Illustrative of this notion, in study 2, we examined in the causal impact of mealtime food photography, which is one form of phone use that has come under scrutiny in both clinical and lay circles. Relative to nonfood photography, food photography disrupted the increase in food consumption that is typically observed across a day [38-40]. Thus, when participants took photographs of their food during a meal, they failed to discriminate between breakfast, lunch, and dinner portion sizes. This pattern is reminiscent of other situational factors (eg, multitasking, eating with peers, and use of digital screens) that distract an individual from eating, such that portion sizes become less driven by internal biological cues (eg, circadian rhythms, homeostatic signals) than by the external environment [46-48]. In turn, this form of external eating may predispose an individual to weight gain over time [47,49,50]. From a clinical perspective, our findings thus support the exploration of mealtime environments, including mealtime phone use habits, when addressing issues of abnormal weight and eating behavior. Nonetheless, we emphasize the preliminary nature of our findings, and the need for further replication.

More broadly, our finding that mealtime phone use is overwhelmingly social highlights another mechanism through which mobile phones may influence eating. This is particularly notable given that participants reported using more nonsocial functions outside the meal context, and because this manner of phone use distinguishes mobile phones from traditional digital screens (where little social interaction occurs). In previous research, one of the most robust determinants of food intake has been found to be the company one eats with: compared to eating alone, eating with others has been repeatedly found to increase consumption, a phenomenon termed *social facilitation* [51-53]. If facilitation can occur with virtual company [54-56], then social interactions during mealtime phone use (eg, through sending text messages) may likewise increase portion sizes [26]. Further research is needed to examine how these social forms of phone use may influence eating behaviors.

Finally, although the discussion has focused on what a phone user does, we also examined how much an individual engages in mealtime phone use (study 1). Here, we found no evidence that frequency altered the risk of obesity: BMIs, routine physical activities, and habitual eating behaviors did not differ significantly between participants who recorded high versus low engagement in mealtime phone use.

Limitations

In presenting these various findings, we note several limitations in both variable and participant selection. First, given the current guidelines on mealtime screen use [12-14], we focused only on phone use within this context. Accordingly, we could not assess the nature and impact of general phone use across the day. Second, we chose to study young adults (aged 18-30 years), the age group most likely to own and be dependent on smartphones [57]. Although this maximized our ability to capture mealtime phone use, further research will need to examine whether our findings generalize to other age groups. Finally, we only collected data on participants' 3 main meals and excluded snacks consumed outside of these times. As meal times are fairly stable, this strategy increased our chance of capturing when participants had eaten (study 1) or were about to eat (study 2). However, future studies can extend our work by exploring the impact of phone use on food intake across the day.

Conclusions

For the first time, we characterized the nature of mealtime phone use and its implications for eating behaviors. By using experience sampling, we tracked a large number of meals in situ (1780 meals across 2 studies) and captured phone use as it occurred in its natural environment [58]. Through this approach, we circumvented recollection failures associated with self-reported phone use data [34] and documented the wide variety of forms through which individuals use their phones during a meal. Further, our preliminary findings suggest that certain types of phone use behaviors may influence food intake. Given the relevance of this topic for clinical practice and guidelines, our findings underscore the importance of follow-up research and the need to move beyond the broad notion of "screen time" to examine how individual phone functions contribute to an obesogenic environment.

Acknowledgments

The authors gratefully acknowledge Alex Meyer for his assistance with the Facebook Messenger bot.

Authors' Contributions

JYYY, JCJL, and EMWT designed the research; JYYY conducted the research; JYYY, JCJL, and EMWT analyzed the data; JYYY and JCJL wrote the paper; and JCJL had primary responsibility for final content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flow diagram of participant selection and dropout.
[PDF File (Adobe PDF File), 75 KB - mhealth_v9i5e22929_app1.pdf]

Multimedia Appendix 2
Demographic, baseline mealtime and phone use, and experience sampling questionnaires.
[PDF File (Adobe PDF File), 33 KB - mhealth_v9i5e22929_app2.pdf]

Multimedia Appendix 3
Validation of measures.
[PDF File (Adobe PDF File), 136 KB - mhealth_v9i5e22929_app3.pdf]

Multimedia Appendix 4
Sample images in the food and nonfood photography conditions.
[PDF File (Adobe PDF File), 2715 KB - mhealth_v9i5e22929_app4.pdf]

Multimedia Appendix 5
Scatterplot of mealtime phone use frequency against number of meals captured.
[PDF File (Adobe PDF File), 149 KB - mhealth_v9i5e22929_app5.pdf]

Multimedia Appendix 6
Multilevel model of food eaten and enjoyment of food (study 2).
[PDF File (Adobe PDF File), 29 KB - mhealth_v9i5e22929_app6.pdf]

References

1. Banks E, Jorm L, Rogers K, Clements M, Bauman A. Screen-time, obesity, ageing and disability: findings from 91 266 participants in the 45 and Up Study. Public Health Nutr 2010 Apr 22;14(1):34-43. [doi: 10.1017/s1368980010000674]
2. Sigmund E, Sigmundová D, Badura P, Kalman M, Hamrik Z, Pavelka J. Sigmund E, Sigmundová D, Badura P, Kalman M, Hamrik Z, Pavelka J. Temporal trends in overweight and obesity, physical activity and screen time among Czech adolescents from 2002 to 2014: a national health behaviour in school-aged children study. Int J Environ Res Public Health 2015 Sep 18;12(9):11848-11868 [FREE Full text] [doi: 10.3390/ijerph120911848] [Medline: 26393638]
3. Duncan MJ, Vandelanotte C, Caperchione C, Hanley C, Mummery WK. Temporal trends in and relationships between screen time, physical activity, overweight and obesity. BMC Public Health 2012 Dec 08;12(1):1060 [FREE Full text] [doi: 10.1186/1471-2458-12-1060] [Medline: 23216917]
4. Staiano AE, Harrington DM, Broyles ST, Gupta AK, Katzmarzyk PT. Television, adiposity, and cardiometabolic risk in children and adolescents. Am J Prev Med 2013 Jan;44(1):40-47 [FREE Full text] [doi: 10.1016/j.amepre.2012.09.049] [Medline: 23253648]
5. Hands BP, Chivers PT, Parker HE, Beilin L, Kendall G, Larkin D. The associations between physical activity, screen time and weight from 6 to 14 yrs: the Raine Study. J Sci Med Sport 2011 Sep;14(5):397-403. [doi: 10.1016/j.jsams.2011.03.011] [Medline: 21531620]
6. Fang K, Mu M, Liu K, He Y. Screen time and childhood overweight/obesity: a systematic review and meta-analysis. Child Care Health Dev 2019 Sep 24;45(5):744-753. [doi: 10.1111/cch.12701] [Medline: 31270831]
7. Blass EM, Anderson DR, Kirkorian HL, Pempek TA, Price I, Koleini MF. On the road to obesity: television viewing increases intake of high-density foods. Physiol Behav 2006 Jul 30;88(4-5):597-604. [doi: 10.1016/j.physbeh.2006.05.035] [Medline: 16822530]
8. Braude L, Stevenson RJ. Watching television while eating increases energy intake. Examining the mechanisms in female participants. Appetite 2014 May;76:9-16. [doi: 10.1016/j.appet.2014.01.005] [Medline: 24462489]

9. Oldham-Cooper R, Hardman C, Nicoll C, Rogers P, Brunstrom J. Playing a computer game during lunch affects fullness, memory for lunch, and later snack intake. Am J Clin Nutr 2011 Mar;93(2):308-313. [doi: 10.3945/ajcn.110.004580] [Medline: 21147857]
10. Moray J, Fu A, Brill K, Mayoral MS. Viewing television while eating impairs the ability to accurately estimate total amount of food consumed. Bariatric Nursing and Surgical Patient Care 2007 Mar;2(1):71-76. [doi: 10.1089/bar.2006.9991]
11. Bilman E, van Kleef E, van Trijp H. External cues challenging the internal appetite control system-Overview and practical implications. Crit Rev Food Sci Nutr 2017 Sep 02;57(13):2825-2834. [doi: 10.1080/10408398.2015.1073140] [Medline: 26463499]
12. Shifrin D, Brown A, Hill D, Jana L, Flinn S. Growing up digital: Media research symposium. : American Academy of Pediatrics; 2015 May 02 Presented at: Growing up digital: Media research symposium; 2015 May 02; Rosemont, Illinois URL: https://www.aap.org/en-us/documents/digital_media_symposium_proceedings.pdf
13. Brown A, Shifrin DL, Hill DL. Beyond 'turn it off': How to advise families on media use. AAP News 2015 Oct;36(10):1. [doi: 10.1542/aapnews.20153610-54]
14. Blum-Ross A, Livingstone S. Families and screen time: Current advice and emerging research. Media Policy Brief 17. In: LSE Media Policy Project. London: London School of Economics and Political Science; 2016.
15. Yen C, Hsiao RC, Ko C, Yen J, Huang C, Liu S, et al. The relationships between body mass index and television viewing, internet use and cellular phone use: the moderating effects of socio-demographic characteristics and exercise. Int J Eat Disord 2010 Sep 13;43(6):565-571. [doi: 10.1002/eat.20683] [Medline: 19718665]
16. Lajunen H, Keski-Rahkonen A, Pulkkinen L, Rose RJ, Rissanen A, Kaprio J. Are computer and cell phone use associated with body mass index and overweight? A population study among twin adolescents. BMC Public Health 2007 Mar 26;7(1):24 [FREE Full text] [doi: 10.1186/1471-2458-7-24] [Medline: 17324280]
17. Fulkerson JA, Loth K, Bruening M, Berge J, Eisenberg ME, Neumark-Sztainer D. Time 2 tlk 2nite: use of electronic media by adolescents during family meals and associations with demographic characteristics, family characteristics, and foods served. J Acad Nutr Diet 2014 Jul;114(7):1053-1058 [FREE Full text] [doi: 10.1016/j.jand.2013.10.015] [Medline: 24361006]
18. Nutrisystem. 1 in 3 Americans can't eat a meal without being on their phone.: Nutrisystem; 2018 Jan 23. URL: https://newsroom.nutrisystem.com/1-in-3-americans-cant-eat-a-meal-without-being-on-their-phone/ [accessed 2021-04-30]
19. Radesky JS, Kistin CJ, Zuckerman B, Nitzberg K, Gross J, Kaplan-Sanoff M, et al. Patterns of mobile device use by caregivers and children during meals in fast food restaurants. Pediatrics 2014 Apr 10;133(4):e843-e849. [doi: 10.1542/peds.2013-3703] [Medline: 24616357]
20. Lee P, Casey M, Wigginton C, Calugar-Pop C. Deloitte's 2019 global mobile consumer survey: Tracking consumers' digital behavior around the world.: Deloitte Insights; 2019. URL: https://www2.deloitte.com/content/dam/insights/us/articles/glob43115_2019-global-mobile-survey/DI_2019-global-mobile-survey.pdf [accessed 2021-04-30]
21. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally.: Pew Research Center; 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-04-30]
22. Barkley JE, Lepp A. Mobile phone use among college students is a sedentary leisure behavior which may interfere with exercise. Computers in Human Behavior 2016 Mar;56:29-33. [doi: 10.1016/j.chb.2015.11.001]
23. Lepp A, Barkley JE, Sanders GJ, Rebold M, Gates P. The relationship between cell phone use, physical and sedentary activity, and cardiorespiratory fitness in a sample of U.S. college students. Int J Behav Nutr Phys Act 2013 Jun 21;10(1):79 [FREE Full text] [doi: 10.1186/1479-5868-10-79] [Medline: 23800133]
24. Stavrinos D, Byington KW, Schwebel DC. Effect of cell phone distraction on pediatric pedestrian injury risk. Pediatrics 2009 Mar 26;123(2):e179-e185. [doi: 10.1542/peds.2008-1382] [Medline: 19171568]
25. Chen Q, Yan Z. Does multitasking with mobile phones affect learning? A review. Computers in Human Behavior 2016 Jan;54:34-42. [doi: 10.1016/j.chb.2015.07.047]
26. Teo E, Goh D, Vijayakumar KM, Liu JCJ. To message or browse? Exploring the impact of phone use patterns on male adolescents? consumption of palatable snacks. Front Psychol 2017 Jan 8;8:2298 [FREE Full text] [doi: 10.3389/fpsyg.2017.02298] [Medline: 29358925]
27. Stapinski H. Restaurants turn camera shy. The New York Times. 2013 Jan 22. URL: https://www.nytimes.com/2013/01/23/dining/restaurants-turn-camera-shy.html [accessed 2021-04-30]
28. Taylor VH. Food Fetish: Society's Complicated Relationship with Food. In: Canadian Journal of Diabetes. 2013 May 01 Presented at: The 3rd National Obesity Summit; 2013 May 01; Vancouver, British Columbia p. S221. [doi: 10.1016/j.jcjd.2013.03.050]
29. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. Prev Med 1999 Dec;29:563-570. [doi: 10.1006/pmed.1999.0585] [Medline: 10600438]
30. Shiffman S, Stone AA. Introduction to the special section: ecological momentary assessment in health psychology. Health Psychology 1998 Jan;17(1):3-5. [doi: 10.1037/h0092706]
31. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. J Youth Adolescence 1977 Sep;6(3):281-294. [doi: 10.1007/bf02138940]

32. Verhagen SJW, Hasmi L, Drukker M, van Os J, Delespaul PAEG. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016 Aug 21;19(3):86-89 [FREE Full text] [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]
33. Smith A. A "week in the life" analysis of smartphone users.: Pew Research Center; 2015 Apr 01. URL: https://www.pewresearch.org/internet/2015/04/01/chapter-three-a-week-in-the-life-analysis-of-smartphone-users/ [accessed 2021-04-30]
34. Andrews S, Ellis DA, Shaw H, Piwek L. Beyond self-report: tools to compare estimated and real-world smartphone use. PLoS One 2015 Oct 28;10(10):e0139004 [FREE Full text] [doi: 10.1371/journal.pone.0139004] [Medline: 26509895]
35. van Strien T, Frijters JER, Bergers GPA, Defares PB. The Dutch Eating Behavior Questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. Int. J. Eat. Disord 1986 Feb;5(2):295-315. [doi: 10.1002/1098-108x(198602)5:2<295::aid-eat2260050209>3.0.co;2-t]
36. Vohs KD, Wang Y, Gino F, Norton MI. Rituals enhance consumption. Psychol Sci 2013 Sep 17;24(9):1714-1721. [doi: 10.1177/0956797613478949] [Medline: 23863754]
37. Lu J, Huet C, Dubé L. Emotional reinforcement as a protective factor for healthy eating in home settings. Am J Clin Nutr 2011 Jul;94(1):254-261. [doi: 10.3945/ajcn.110.006361] [Medline: 21613564]
38. Schusdziarra V, Hausmann M, Wittke C, Mittermeier J, Kellner M, Naumann A, et al. Impact of breakfast on daily energy intake--an analysis of absolute versus relative breakfast calories. Nutr J 2011 Jan 17;10(1):5 [FREE Full text] [doi: 10.1186/1475-2891-10-5] [Medline: 21241465]
39. de Castro JM. Circadian rhythms of the spontaneous meal pattern, macronutrient intake, and mood of humans. Physiology & Behavior 1987 Jan;40(4):437-446. [doi: 10.1016/0031-9384(87)90028-x]
40. de Castro JM. Heritability of diurnal changes in food intake in free-living humans. Nutrition 2001 Sep;17(9):713-720. [doi: 10.1016/s0899-9007(01)00611-6]
41. Halford JC, Gillespie J, Brown V, Pontin EE, Dovey TM. Effect of television advertisements for foods on food consumption in children. Appetite 2004 Apr;42(2):221-225. [doi: 10.1016/j.appet.2003.11.006] [Medline: 15010186]
42. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. The Lancet 2004 Jan;363(9403):157-163. [doi: 10.1016/s0140-6736(03)15268-3]
43. Department of Statistics. Population trends 2020. Singapore: Department of Statistics, Ministry of Trade & Industry URL: https://www.singstat.gov.sg/-/media/files/publications/population/population2020.pdf [accessed 2021-04-30]
44. Smith A. Americans and their cell phones.: Pew Research Center; 2011 Aug 15. URL: https://www.pewresearch.org/internet/2011/08/15/americans-and-their-cell-phones/ [accessed 2021-04-30]
45. Silver L, Smith A, Johnson C, Taylor K, Jiang J, Anderson M, et al. Mobile connectivity in emerging economies.: Pew Research Center; 2019 Mar 07. URL: https://www.pewresearch.org/internet/2019/03/07/mobile-connectivity-in-emerging-economies/ [accessed 2021-04-30]
46. Long S, Meyer C, Leung N, Wallis DJ. Effects of distraction and focused attention on actual and perceived food intake in females with non-clinical eating psychopathology. Appetite 2011 Apr;56(2):350-356. [doi: 10.1016/j.appet.2010.12.018] [Medline: 21185894]
47. Hetherington MM. Cues to overeat: psychological factors influencing overconsumption. Proc. Nutr. Soc 2007 Feb 28;66(1):113-123. [doi: 10.1017/s0029665107005344]
48. Brunstrom JM, Mitchell GL. Effects of distraction on the development of satiety. Br J Nutr 2006 Oct;96(4):761-769. [Medline: 17010237]
49. Schachter S. Obesity and eating. Internal and external cues differentially affect the eating behavior of obese and normal subjects. Science 1968 Aug 23;161(3843):751-756. [doi: 10.1126/science.161.3843.751] [Medline: 5663800]
50. Herman C, Polivy J. External cues in the control of food intake in humans: the sensory-normative distinction. Physiol Behav 2008 Aug 06;94(5):722-728. [doi: 10.1016/j.physbeh.2008.04.014] [Medline: 18499202]
51. De Castro JM. Socio-cultural determinants of meal size and frequency. Br J Nutr 1997 Apr 09;77 Suppl 1(S1):S39-54; discussion S54. [doi: 10.1079/bjn19970103] [Medline: 9155493]
52. Herman CP. The social facilitation of eating. A review. Appetite 2015 Mar;86:61-73. [doi: 10.1016/j.appet.2014.09.016] [Medline: 25265153]
53. Herman CP, Roth DA, Polivy J. Effects of the presence of others on food intake: a normative interpretation. Psychol Bull 2003 Nov;129(6):873-886. [doi: 10.1037/0033-2909.129.6.873] [Medline: 14599286]
54. Park S, Catrambone R. Social facilitation effects of virtual humans. Hum Factors 2016 Jul 02;49(6):1054-1060. [doi: 10.1518/001872007x249910]
55. Snyder A, Anderson-Hanley C, Arciero P. Virtual and live social facilitation while exergaming: competitiveness moderates exercise intensity. J Sport Exerc Psychol 2012 Apr;34(2):252-259. [doi: 10.1123/jsep.34.2.252] [Medline: 22605365]
56. Anderson-Hanley C, Arciero P, Snyder. Social facilitation in virtual reality-enhanced exercise: competitiveness moderates exercise effort of older adults. Clin Interv Aging 2011 Oct;2011(6):275-280. [doi: 10.2147/cia.s25337]
57. Mobile fact sheet.: Pew Research Center; 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2019-12-30]
58. Scollon CN, Kim-Prieto C, Scollon CN. Experience sampling: promises and pitfalls, strengths and weaknesses. Journal of Happiness Studies 2003;4(1):5-34. [doi: 10.1023/a:1023605205115]

Abbreviations

DEBQ: Dutch Eating Behavior Questionnaire

Edited by G Eysenbach; submitted 02.08.20; peer-reviewed by A Davies, L Brennan; comments to author 10.09.20; revised version received 30.09.20; accepted 11.04.21; published 06.05.21.

Please cite as:
Yong JYY, Tong EMW, Liu JCJ
Meal-time Smartphone Use in an Obesogenic Environment: Two Longitudinal Observational Studies
JMIR Mhealth Uhealth 2021;9(5):e22929
URL: https://mhealth.jmir.org/2021/5/e22929
doi:10.2196/22929
PMID:33955842

Original Paper

Real-Time UV Measurement With a Sun Protection System for Warning Young Adults About Sunburn: Prospective Cohort Study

June K Robinson[1], MD; Shiv Patel[1], BA; Seung Yun Heo[2], BSc; Elizabeth Gray[3], MSc; Jaeman Lim[4], BSc; Kyeongha Kwon[2,5], PhD; Zach Christiansen[4], BSc; Jeffrey Model[4], BSc; Jacob Trueb[2], MCS; Anthony Banks[2], BSc; Mary Kwasny[3], ScD; John A Rogers[2], PhD

[1]Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Department of Biomedical Engineering, Center for Bio-Integrated Electronics, Simpson Querrey Institute for BioNanotechnology, Northwestern University, Evanston, IL, United States

[3]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[4]Wearifi Inc, Evanston, IL, United States

[5]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Corresponding Author:
June K Robinson, MD
Northwestern University Feinberg School of Medicine
132 East Delaware Place, #5806
Chicago, IL, 60611
United States
Phone: 1 3129433703
Email: june-robinson@northwestern.edu

Abstract

Background: Melanoma is attributable to predisposing phenotypical factors, such as skin that easily sunburns and unprotected exposure to carcinogenic UV radiation. Reducing the proportion of young adults who get sunburned may reduce the incidence of melanoma, a deadly form of skin cancer. Advances in technology have enabled the delivery of real-time UV light exposure and content-relevant health interventions.

Objective: This study aims to examine the feasibility of young adults performing the following tasks daily: wearing a UV dosimeter, receiving text messages and real-time UV-B doses on their smartphone, and responding to daily web-based surveys about sunburn and sun protection.

Methods: Young adults aged 18-39 years (n=42) were recruited in the United States in June 2020 via social media. Participants received the UV Guard sun protection system, which consisted of a UV dosimeter and a smartphone app. During 3 consecutive periods, intervention intensity increased as follows: real-time UV-B dose; UV-B dose and daily behavioral facilitation text messages; and UV-B dose, goal setting, and daily text messages to support self-efficacy and self-regulation. Data were self-reported through daily web-based surveys for 28 days, and UV-B doses were transmitted to cloud-based storage.

Results: Patients' median age was 22 years (IQR 20, 29), and all patients had sun-sensitive skin. Sunburns were experienced during the study by fewer subjects (n=18) than those in the preceding 28 days (n=30). In July and August, the face was the most commonly sunburned area among 13 body locations; 52% (22/42) of sunburns occurred before the study and 45% (19/42) occurred during the study. The mean daily UV-B dose decreased during the 3 periods; however, this was not statistically significant. Young adults were most often exercising outdoors from 2 to 6 PM, walking from 10 AM to 6 PM, and relaxing from noon to 2 PM. Sunburn was most often experienced during exercise (odds ratio [OR] 5.65, 95% CI 1.60-6.10) and relaxation (OR 3.69, 95% CI 1.03-4.67) relative to those that did not exercise or relax in each category. The self-reported exit survey indicated that participants felt that they spent less time outdoors this summer compared to the last summer because of the COVID-19 pandemic and work. In addition, 38% (16/42) of the participants changed their use of sun protection based on their app-reported UV exposure, and 48% (20/42) shifted the time they went outside to periods with less-intense UV exposure. A total of 79% (33/42) of the participants were willing to continue using the UV Guard system outside of a research setting.

Conclusions: In this proof-of-concept research, young adults demonstrated that they used the UV Guard system; however, optimization was needed. Although some sun protection behaviors changed, sunburn was not prevented in all participants, especially during outdoor exercise.

Trial Registration: ClinicalTrials.gov NCT03344796; http://clinicaltrials.gov/ct2/show/NCT03344796

(JMIR Mhealth Uhealth 2021;9(5):e25895) doi:10.2196/25895

KEYWORDS

sun protection; UV dosimeter; health promotion technology; melanoma; sunburn; preventive medicine; mobile phone

Introduction

Background

Young adults aged between 18 and 39 years often engage in activities that expose them to high amounts of carcinogenic UV radiation [1]. Melanoma, the second most diagnosed cancer in young adults [2], is attributable to both predisposing genetic and phenotypical factors, such as having skin that easily gets sunburned, as well as unprotected exposure to UV. Leisure-time physical activity, such as sports engaged in by young adults to maintain fitness and socialize, has been found to be associated with an increased risk of melanoma [3]. Outdoor leisure-time physical activity is frequently performed with limited skin coverage by clothing and is associated with an increased risk of sunburn [4]. In contrast to the prior focus on the importance of childhood sunburns, growing epidemiological evidence indicates that sunburns sustained during all life periods, including adulthood, increase the risk of melanoma [5]. In the United States, 50.2% of young adults aged 18-29 years and 51.2% of individuals with sun-sensitive skin experience sunburn each year [6]. Thus, limiting the amount of unprotected UV exposure during young adulthood has the potential to decrease the risk of melanoma, especially among high-risk individuals. Various sun protection behaviors, such as avoiding the sun during peak UV hours (10 AM-4 PM), wearing protective clothing, and applying broad-spectrum sunscreen with a sun protection factor (SPF) of 30 or higher, are recommended.

A limited number of studies have evaluated the impact of sun protection interventions on self-reported sun protection usage by at-risk young adults. If the biological outcome of sunburn was collected in these studies, participants were asked to recall the number of sunburns in the last month, which was subject to recall bias. The use of a web-based intervention was effective in reducing UV exposure and increasing skin protection behaviors over 12 weeks [7]. In a 3-month Australian study, individuals equipped with personal UV monitors improved their sun protection behaviors on weekends compared with a control group [8]. The UV monitor provided the personal real-time UV dose; however, the cumulative UV daily dose was not determined. In addition, users sought more control over the feedback method provided by the UV monitor, which sounded an alarm at predefined UV thresholds based on anticipated sunburn associated with skin types. Specifically, participants noted that the ability to tailor the alert as a custom alarm or a subtle vibration would be more appealing. Within the same study, another group of participants received a different intervention involving sun protection notifications delivered through a mobile app. However, many users of the app reported that the notifications were repetitive and offered information that was already known.

Objective

Findings from our past research informed the development of the sun protection mobile app, wearable UV sensor, UV Guard program, and text messages used in this study [9-11]. The wearable UV sensor collected and reported UV-B measurements in real time through an accompanying mobile app on the user's smartphone. These UV measurements were also stored in a cloud database of personal UV exposure and were accessed daily. The purpose of this study is to evaluate the feasibility of young adult participants wearing the UV sensor each day, receiving daily sun protection messages by SMS texts, and reporting their sun exposure and protection behaviors on a web-based app daily for a span of 28 days. Sun exposure and protection and the risk of sunburn among a cohort of young adults were assessed.

Methods

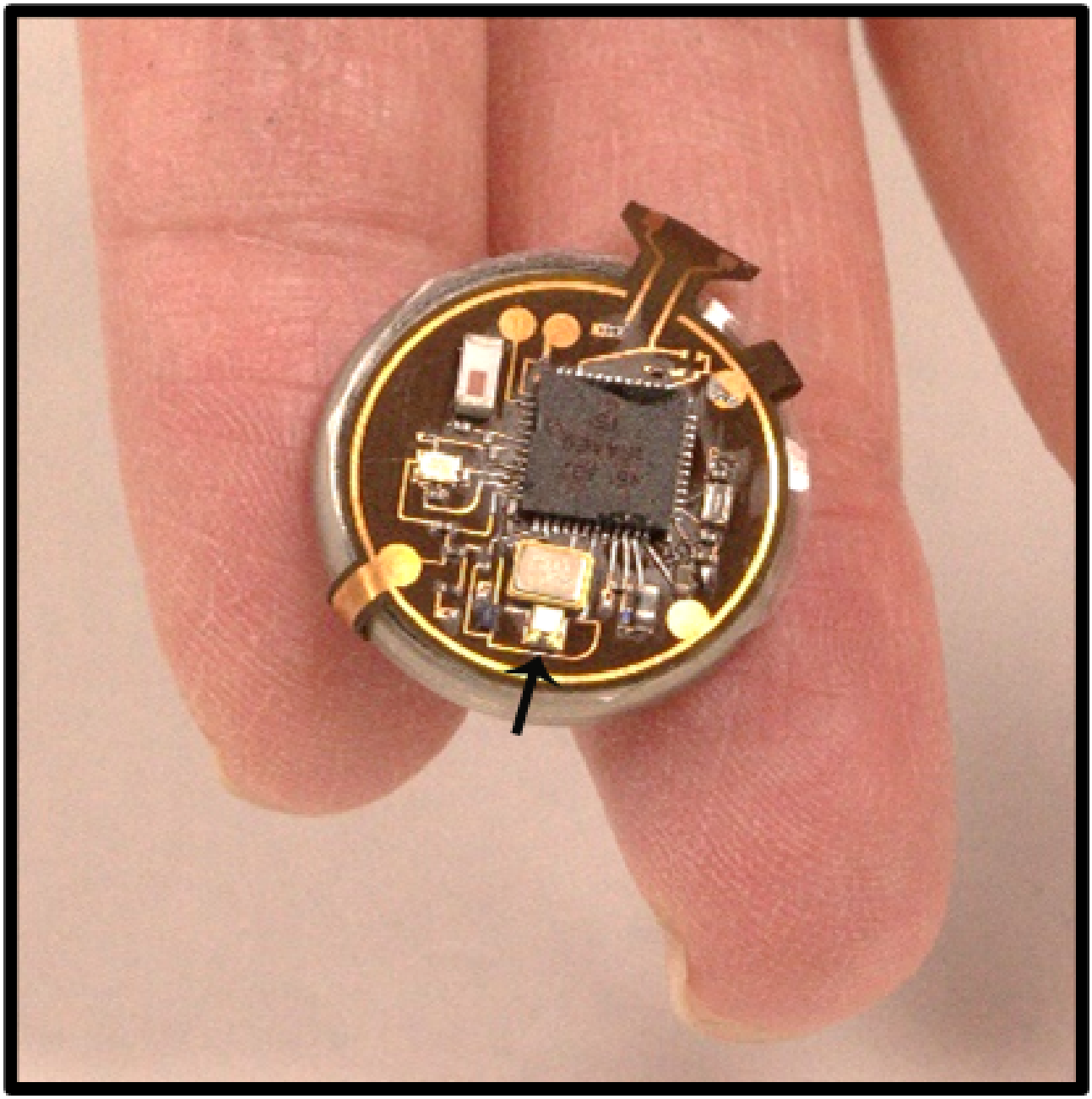
Recruitment

In June 2020, participants were recruited by posting electronic announcements on college websites and high school alumnae organizations in the Midwestern and Southeastern United States. The announcement stated that the aim of the study was to provide real-time UV exposure to participants to prevent sunburn. Eligibility criteria were age 18-39 years, having skin that gets pink (only just perceptible reddening of the skin) after being in the sun, normally spending at least 30 consecutive minutes a day outdoors, having a home address with a direct mail address to receive the UV dosimeter, having reliable internet access, and having a smartphone with iOS version 13.0 or above that was able to support the app, be willing to wear the sensor on the wrist similar to a watch for 28 days, and complete daily web-based surveys. Young adults, who were interested in participating in the study, clicked a link provided, after which they were directed to a survey website (REDCap [Research Electronic Data Capture]) [12].

Intervention Device

The directions for using the wireless miniaturized UV dosimeter with a Bluetooth connection to the smartphone and cloud-based storage of the daily UV-B dose were provided by email, and printed directions were enclosed with the device [13,14]. The device was shipped to the participants by Federal Express (Figure 1). After receiving the device, participants logged onto the web-based survey [12] to complete the REDCap survey. The Institutional Review Board of Northwestern University approved the study, and written informed consent was obtained from each participant (Multimedia Appendix 1). For people who qualified to participate in this study, an electronic gift card was sent by email at the completion of the study.

Figure 1. Wireless UV-B dosimeter worn on the wrist (arrow indicates a UV-B photodiode).



The ultra–low-power digital dosimeter platform used in this study provided continuous real-time UV-B doses wirelessly to the users' smartphone, with immediate access to the dose received in relation to their sunburn threshold. The operation relies on a UV-B photodiode that continuously accumulates charge on a storage capacitor such that the resulting voltage corresponds directly to the exposure dose via a calibration factor. The use of this accumulation detection module with an advanced, light-adaptive electronic control circuit enabled exceptionally high levels of power efficiency for a long battery life [13]. When the exposure dose exceeded a predefined threshold value, the device automatically and wirelessly transmitted the dose value to a smartphone. For field deployment, devices such as those shown in Figure 1 were packaged into housings designed for digital watches for mounting on the wrist. This type of wearable dosimeter, with a smartphone and a cloud computing system, was used to support personalized UV-B monitoring (Figure 1).

Study Design: Text Messages

In June and July 2019, semistructured interviews were conducted with 30 young adults (15 males and 15 females) to develop text messages using the principles of Social Cognitive Theory (self-efficacy and outcome expectancies) and qualitative data analytic methods previously reported [15,16]. As young adults wanted knowledge (sun protection tips) from an expert, a series of 15 knowledge-based messages to facilitate sun protection behavior were assessed by Likert scale (range 1-5; 1=very useful; 5=not useful) and refined to 8 items with a mean score of 1.8 (SD 0.99). Five outcome expectancy items were refined to 2 items, with a mean score of 1.6 (SD 0.75). Four self-regulation items had a mean score of 2.6 (SD 1.3).

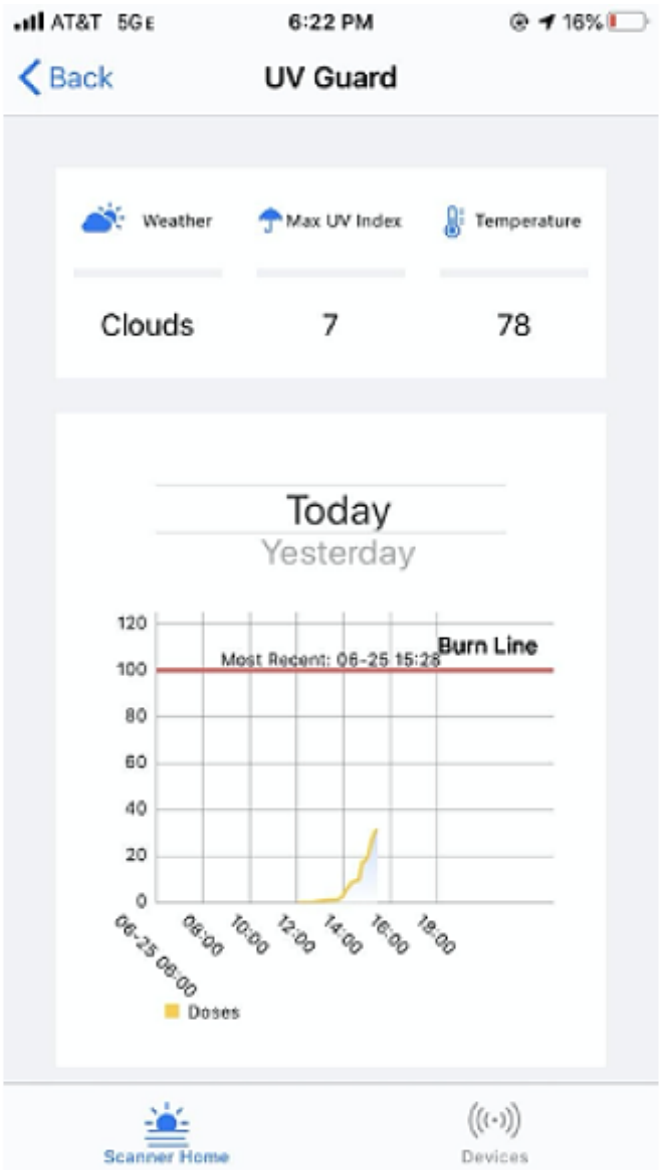
Self-efficacy items, which were not tailored to the individual behavior of users, had a mean score of 2.8 (SD 1.4).

Participants did not report if they received or read the text messages.

Study Design: Interventional Study

From July to August 2020, an interventional study with a series of components was conducted over 3 periods. Initially, participants selected their skin type based on their perceived sun sensitivity to sunburn, which corresponded to skin type I, II, or III. The minimal erythema dose (MED) for skin type I, II, and III is 200 J/m^2, 250 J/m^2, and 300 J/m^2, respectively [17]. Thus, a sunburn line was established for each participant according to their self-reported skin type (Figure 2). Sunburn was defined as only just perceptible reddening of the skin (pink) at 24 hours after UV exposure [18]. Intermittent UV exposure may produce an MED equivalent to a single UV exposure, which supports the cautious approach of using cumulative UV daily exposure to warn participants of impending sunburn.

Figure 2. Smartphone screen showing the real-time UV-B dose in relation to the threshold dose at which the participant would get a sunburn; the predicted weather, including cloud cover; UV index; and temperature.



The study periods were as follows: (1) days 0-7: participants reviewed the real-time daily cumulative UV-B exposure provided on the screen of their smartphone in a graph showing the UV-B dose at which their skin would sustain a sunburn displayed on the graph with a red line (Figure 2); (2) days 8-17: UV-B dose plus daily text messages at 11 AM to enhance sun protection (Textbox 1); and (3) days 18-28: on day 18, each participant selected primary and secondary goals to improve their sun protection and indicated their intention to continue performing the sun protection. In addition to the UV-B dose, participants received text messages described in Textbox 1 whose content shifted from behavioral facilitation (knowledge) and outcome expectancy in the first 16 days to self-efficacy and self-regulation following goal setting on day 18. Young adults selected their preferred time to receive text messages.

Textbox 1. Daily text messages.

Days 0-7

- No text message

Day 8

- *Remember, the purpose of sun protection is to prevent a skin cancer and keep your skin from aging. Make sure to use sun protection; sunscreen, wearing sun protective clothing, and stay in the shade when possible!* [Behavioral facilitation]

Day 9

- *Sun peak intensity alert! Did you know UV rays, which are the harmful rays of the sun, are strongest from 10 AM to 4 PM? Be extra careful during this time.* [Behavioral facilitation]

Day 10

- *Try self-tanning lotion to safely appear tanned.* [Behavioral facilitation]

Day 11

- *An SPF of 30 indicates protection against 97% of harmful UV rays. Apply sunscreen 15-20 min before going outside and re-apply if you're still in the sun after two hours.* [Outcome expectancy]

Day 12

- *Lips need love too! Be sure to apply a lip balm, Chap Stick, and/or lipstick with an SPF of at least 30 to your lips. Lips can get a melanoma too.* [Behavioral facilitation]

Day 13

- *It is easy to miss spots when you are applying sunscreen. Using 2 coats helps miss fewer areas! Do one area of the body, let it dry, and then reapply another coat.* [Behavioral facilitation]

Day 14

- *Sunburns can happen on cloudy days. More than 80% of UV rays can pass through clouds. UV rays also reflect off water and concrete.* [Outcome expectancy]

Day 15

- *When shopping for a sunscreen, be sure to select one that says it is broad-spectrum, has an SPF of >30 and is water resistant. Be sure to check the expiration dates of bottles you already have.* [Behavioral facilitation]

Day 16

- *As you sweat, your sunscreen washes away. Be sure to reapply your sunscreen after 80 min. The new sunscreen layer is effective when the old layer is used up at 2 hours. This applies to outdoor activities like exercising, errands, and commuting!* [Behavioral facilitation]

Day 17

- *Keeping track of your UV exposure each day will get you a little closer to achieving your sun protection goals!* [Self-efficacy]

Day 18

- *Pay it forward! Text or call a family member and friend and remind them of the importance of sun protection!* [Self-efficacy]

Day 18 Goal setting

- *How will you meet your goal to engage in sun-protected outdoor activity? (Select your first and second choices) Tomorrow, I will:*
  - *Apply sunscreen to all areas of my body that may be exposed to the sun.*
  - *Apply sunscreen before I go outdoors.*
  - *Wear a hat when I am outdoors.*
  - *Wear a shirt that covers my shoulders when I am outdoors.*
  - *Plan my outdoor activities to avoid being outside from 10 AM to 2 PM.*
  - *Pay attention to the strength of the sun by checking the UV report 15 min after I go outside.*
  - *Be careful not to exceed the amount of UV my skin can tolerate.*

Day 19 Goal intention

- *You picked the goal to [...]: Do you intend to keep doing this?*

Day 19

- *As a pick-me-up, take a few minutes out of your day and go for a walk outdoors. Be sure to use sun protection and walk where there is the most shade!* [Self-regulation]

Day 20

- *Keep sunscreen someplace where you'll see it, like near your toothpaste. Make it part of your morning routine.* [Behavioral facilitation]

Day 21

- *If you forget to apply sunscreen and do not have sun protective clothing or a hat, be sure to utilize shade!* [Self-regulation]

Day 22

- *Two-thirds of adults get a sunburn more than once each year. Pledge to not be one of them.* [Self-regulation]

Day 23

- *A sunburn does not always hurt or blister. If you press on skin that is slightly pink and the color goes away, you have a sunburn.* [Self-regulation]

Day 24

- *Try keeping a hat or sunscreen somewhere that's easy to get when you are on-the-go, such as your car or backpack.* [Behavioral facilitation]

Day 25

- *A good way to figure out if a tree gives enough shade for complete sun protection is to look at the ground and see if spots of sunlight show on the ground.* [Self-regulation]

Day 26

- *Practicing safe sun habits is my lifelong habit.* [Self-efficacy]

Day 27

- *Getting in the habit of protecting your skin now will help you make it a habit for many years to come!* [Self-efficacy]

Day 28

- *Good for you! You met your goal of improving your sun protection.* [Self-efficacy]

Data were collected during July and August 2020. If a participant failed to complete the daily self-reported survey or the UV-B dose was not transferred to the cloud, the participant received an email reminder the next day.

The Institutional Review Board of Northwestern University approved the study protocol. Participants provided written consent and were offered a US $200 gift card after completing the final survey.

Measures

Baseline self-reported responses included age, gender, race or ethnicity, skin type (sun sensitivity), family and personal history of skin cancer, sunburns, and body parts with a sunburn in the last 28 days; knowledge and attitudes about sunburn, sun exposure, and sun protection; estimated hours outside 10 AM and 4 PM on weekdays and weekend days in the past 28 days; and sun protection used in the past 28 days, including wearing sunscreen, wearing a shirt with sleeves or a T-shirt, wearing sunglasses, or staying in the shade (Likert scale: 1=never; 3=sometimes; 5=always]; Table 1). These measures were used in prior sun exposure and protection studies [9,19].

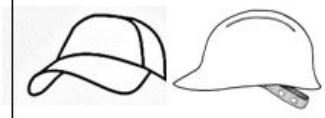
Table 1. Schedule of measures.

Measure	Day 0: baseline	Days 0-7: observe	Days 8-17: daily text messages	Day 18: structured goal setting	Days 18-28: daily text messages
Self-report					
Demographics	✓				
Sunburn in last 28 days	✓				
Confidence and anxiety	✓	D[a]7	D16		D28
Daily sunburn or sun protection		✓	✓	✓	✓
Structured goal				✓	
System Usability Scale					✓
Willingness to continue use					✓
Sensor					
Daily UV	✓	✓	✓	✓	✓
Intervention					
Daily text messages			✓	✓	✓
UV exposure visualization		✓	✓	✓	✓

[a]D: day.

Daily outside activities were elicited for 2-hour blocks of time (6-7:59 AM, 8-9:59 AM, 10-11:59 AM, noon-1:59 PM, 2 to 3:59 PM, and 4 to 5:59 PM). Participants also reported sun protection for the same 2-hour blocks by selecting a picture of the type of clothing worn (Figure 3) and any clothing changes during the day and reported the application of sunscreen, reapplication, and SPF of the sunscreen. Daily sunburns of 13 body parts were self-reported each evening in the web-based survey (Multimedia Appendix 2). The temperature and sun-clear or cloudy or rain-stormy conditions were obtained from the Open Weather Map for the geographic location of the participant for the 2-hour block of time used to self-report sun protection. Daily UV-B exposure for each participant was obtained from the wearable UV sensor.

Figure 3. Example of a self-reported measure for clothing covering the head and the neck. N/A: not applicable.

Coverage Option	Coverage Example Picture	Sun Protection Provided	Sun Protection Not Provided
1. No Coverage		N/A	No sun protection of scalp, face, ears or neck.
2. Visor		Sun protection of upper 1/3 of face	No sun protection of scalp, lower 2/3 of face, ears, or neck
3. Baseball cap/safety hat		Sun protection of scalp and upper 1/3 of face	No sun protection of lower 2/3 of face, ears, and neck (back and front)
4. 4 inch brim hat/bucket hat/sun hat		Sun protection of scalp, most of face, ears, and partial sun protection of neck	No sun protection of front of neck
5. Sun protection hat with neck flap		Sun protection of scalp, upper 1/3 of face, ears, and back of neck	No sun protection for lower half of face and front of neck

UV exposure was transmitted from the personal UV dosimeter to users' smartphone and to the cloud database. Exposure during the preceding 24 hours was downloaded each day.

On day 18, the REDCap system offered participants the opportunity to select 2 items from the following list of possible goals to implement tomorrow: (1) "Apply sunscreen to all of the areas of my body that may be exposed to the sun," (2) "Apply sunscreen before I go outdoors," (3) "Wear a hat when I am outdoors," (4) "Wear a shirt that covers my shoulders when I am outdoors," (5) "Plan my outdoor activities to avoid being outside from 10 AM to 2 PM," (6) "Pay attention to the strength of the sun by checking the UV Guard report 15 minutes after I go outside," and (7) "Be careful not to exceed the amount of UV my skin can tolerate." The following day, participants were reminded of their primary goal choice and asked if they intended to keep doing it.

Anxiety was assessed with self-reported responses to 11 items ranging from 1 (strongly disagree) to 5 (strongly agree) used in previous research (range 11-55) [9]. Similarly, confidence in practicing sun protection was assessed using 11 items (range 11-55).

Statistical Analysis

Demographic characteristics were summarized using medians and IQRs for age and counts and percentages for other demographics, including knowledge scores. Generalized linear mixed models with logit links were used to assess the estimated probabilities of probabilities of reporting daily sunburns for each day, body part, study period, and time of day, and an identity link was used to model the UV dose. Both models assumed an unstructured covariance matrix. Least square means and SDs were presented as the mean percentage of days that participants reported a sunburn within each study period after adjusting for weather (cloudy, rainy, or clear), UV dose, and the use of sun protection. In addition to type 3 tests for the main effects of the study period on reported sunburns, reported activities during the day were also examined for associations. Odds ratios (ORs) and 95% CIs are presented. Descriptive statistics were used to describe the goals and summarize the study experiences of the subjects at a study exit interview. Estimated anxiety and confidence regarding sun protection were compared across time points using repeated measures analysis of variance models and reported using means and SDs. All analyses were run using R 3.6.0 (The R Foundation) at a nominal type I error rate of 5% [20].

Data Exclusion

One UV dosimeter malfunctioned; therefore, 1 participant's data were excluded from the UV-B exposure reported in Table 2.

Table 2. Daily sunburns and UV-B dose experienced by participants and mean proportion of each day spent outside by all participants during the 3 periods (days 0-7, days 8-17, and days 18-28).

Variables	Days 0-7	Days 8-17	Days 18-28	*P* value[a]
Population, n	42	42	42	N/A[b]
Individual participants				
Number of sunburns, n (%)				.55
0	24 (57)	27 (64)	25 (59)	
1	12 (29)	8 (19)	8 (19)	
2	4 (10)	5 (12)	6 (14)	
More than 2	2 (5)	2 (5)	3 (7)	
UV-B daily dose (J/m^2), mean (SD)[c]	91.96 (115.34)	74.97 (82.32)	62.73 (71.44)	.08
Sample				
Proportion of each day spent outside, mean (SD)	0.87 (0.19)	0.84 (0.21)	0.81 (0.24)	.05
6-7:59 AM	0.15 (0.25)	0.14 (0.20)	0.14 (0.21)	.93
8-9:59 AM	0.19 (0.21)	0.18 (0.20)	0.18 (0.22)	.93
10-11:59 AM	0.33 (0.28)	0.31 (0.23)	0.31 (0.22)	.74
Noon-1:59 PM	0.43 (0.29)	0.41 (0.22)	0.36 (0.24)	.20
2-3:59 PM	0.44 (0.26)	0.37 (0.19)	0.39 (0.25)	.24
4-5:59 PM	0.47 (0.27)	0.45 (0.30)	0.48 (0.28)	.67

[a]*P* values are from the main effects of time from repeated measures general linear models (log or identity link).

[b]N/A: not applicable.

[c]The UV device malfunctioned for 1 person; thus, the participant's data were removed.

Results

Participant Characteristics

A total of 44 young adults were enrolled in the study and 42 completed the 28-day study. All participants self-reported having either very sun-sensitive skin or average sun-sensitive skin (Table 3). Within 5 days of enrolling, 2 participants stopped completing the daily self-reported survey and wearing the sensor; therefore, they ceased study participation. All of the remaining 42 participants wore the UV sensor daily and completed a daily self-reported survey for 28 days. The time most frequently selected to receive text messages was 11 AM.

Table 3. Participant characteristics.

Characteristics	Values
Population, n	42
Age (years), median (first quartile, third quartile)	22 (20, 29)
Gender, n (%)	
Female	28 (67)
Male	14 (33)
Race, n (%)	
White	36 (86)
Asian	3 (7)
Other	2 (5)
Prefer not to answer	1 (2)
Ethnicity, n (%)	
Non-Hispanic	39 (93)
Hispanic	3 (7)
Skin type, n (%)	
I. Very sun sensitive	25 (60)
II. Average sun sensitive	17 (40)
III. Low sun sensitive	0 (0)
Family history of skin cancer, n (%)	
No	21 (50)
Yes	21 (50)
Personal history of skin cancer, n (%)	
No	42 (100)
Yes	0 (0)

Baseline Knowledge and History of Sunburn and Sun Protection in the Preceding 28 Days

Participants were asked to identify the characteristics of sunburns among the 6 offered (pink skin, red skin, pain, peeling, blistering, and skin hot to the touch). A total of 50% (22/44) of the participants were able to identify all 6 characteristics (range 1-6), with the most commonly overlooked characteristic being blistering. This was missed by 30% (13/44) of the participants.

Knowledge of sun strength and protection was tested by asking a series of questions regarding when the sun was strongest (time and month), UV rays, and clothing protection. On a scale of 0-6, scores ranged from 2 to 6, with 43% (18/42) of the participants scoring 2-3, 36% (15/42) scoring a 4, and 21% (9/42) scoring 5-6. All participants were able to recognize that the sun was strongest from 10 AM to 4 PM and that the best head protection was a hat with a 4-inch brim and neck flap. In contrast, 66% (29/44) and 84% (37/44) of the participants were able to identify the danger of UV-A and UV-B light, respectively; only 20% (9/44) of the participants recognized that the sun was strong enough to burn as early as March.

Among the 42 participants compared for getting a sunburn preceding and during the study, 71% (30) reported having sunburns in the 28 days preceding the study. Participants reported being outdoors for 31 minutes to 1 hour (26/42, 62%), more than 1 hour and up to 2 hours (11/42, 26%), and more than 2 hours and up to 3 hours (5/42, 12%) in the 28 days preceding the study. Frequency was reported as often or sometimes by 64% (27/42) of participants wearing sunscreen, 52% (22/42) wearing a shirt with sleeves, 21% (9/42) wearing a hat with a brim, and 71% (30/42) seeking shade. Wearing sunglasses was always done by 26% (11/42) and often or sometimes by 35% (15/42).

Characterization of Sunburn and Sun Exposure During the 3 Periods

Sunburns were consistently experienced during all 3 periods. Although the mean UV-B exposure over the 3 periods declined, the differences were not statistically significant (Table 2). The mean proportion of each day spent outdoors by the sample demonstrated a statistically significant change in the 3 periods. During July and August, the ultraviolet index (UVI) was between 7 and 10, depending on the geographic location and local weather conditions [21,22]. This UVI was sufficient to produce sunburns in unprotected skin. Sunburn was most often experienced by participants during exercise (OR 5.65, 95% CI 1.60-6.10) and relaxation (OR 3.69, 95% CI 1.03-4.67) and less often with walking (OR 1.36, 95% CI 0.39-1.07) than by those who did not exercise, walk, or have outdoor relaxation.

The 13 sunburned body regions were compared 28 days before and during the 28 days of the study. Sunburns were experienced during the study by fewer subjects (n=18) than in the 28 days preceding the study (n=30). A statistically significant reduction in sunburn was on the shoulders and chest. (Table 4). The face was the most commonly sunburned area among the 13 body locations, with 52% (22/42) of sunburns before the study and 45% (19/42) during the study. Other body regions with sunburn reduction trends were the scalp and the back.

Table 4. Body region sunburned before and during the study for a sample of 42 participants.

Body regions	Participants sunburned before the study, n (%)	Participants sunburned during the study, n (%)	*P* value[a]
Face	22 (52)	19 (45)	.65
Neck	11 (26)	9 (21)	.68
Ears	4 (10)	3 (7)	.99
Scalp	5 (12)	1 (2)	.13
Shoulders	20 (48)	9 (21)	.01[b]
Back	9 (21)	4 (10)	.23
Chest	14 (33)	6 (14)	.02[b]
Stomach	4 (10)	0 (0)	__[c]
Arms	13 (31)	17 (41)	.45
Hands	3 (7)	5 (12)	.68
Buttocks	1 (2)	1 (2)	.99
Legs	8 (19)	8 (19)	.99
Feet	7 (17)	5 (12)	.77

[a]*P* values are from the McNemar test of paired proportions. As this was an exploratory analysis, no correction for multiple comparisons was made for the type I error rate.

[b]Shoulders and chest had significant decrease in sunburn.

[c]Not available. Owing to a limited sample size, the test could not be performed.

Sun Exposure During Outdoor Activities

Young adults were most often outdoors exercising from 2 to 6 PM, walking from 10 AM to 6 PM, and relaxing from noon to 2 PM (Figures 4-7). Public health strategies during the COVID-19 pandemic in the United States restricted team sports and closed gyms. Shelter-in-place policies asked workers to perform their jobs from home. There was no significant difference in the proportion of time devoted to these outdoor activities during the 3 periods.

In all 3 periods, the proportion of the day spent outside with an unprotected face was greatest from noon to 6 PM, which resulted in facial sunburn (Figure 8).

Figure 4. Mean proportions of each of the days spent outside by the sample for any activity during the 3 periods stratified by 2-hour periods.

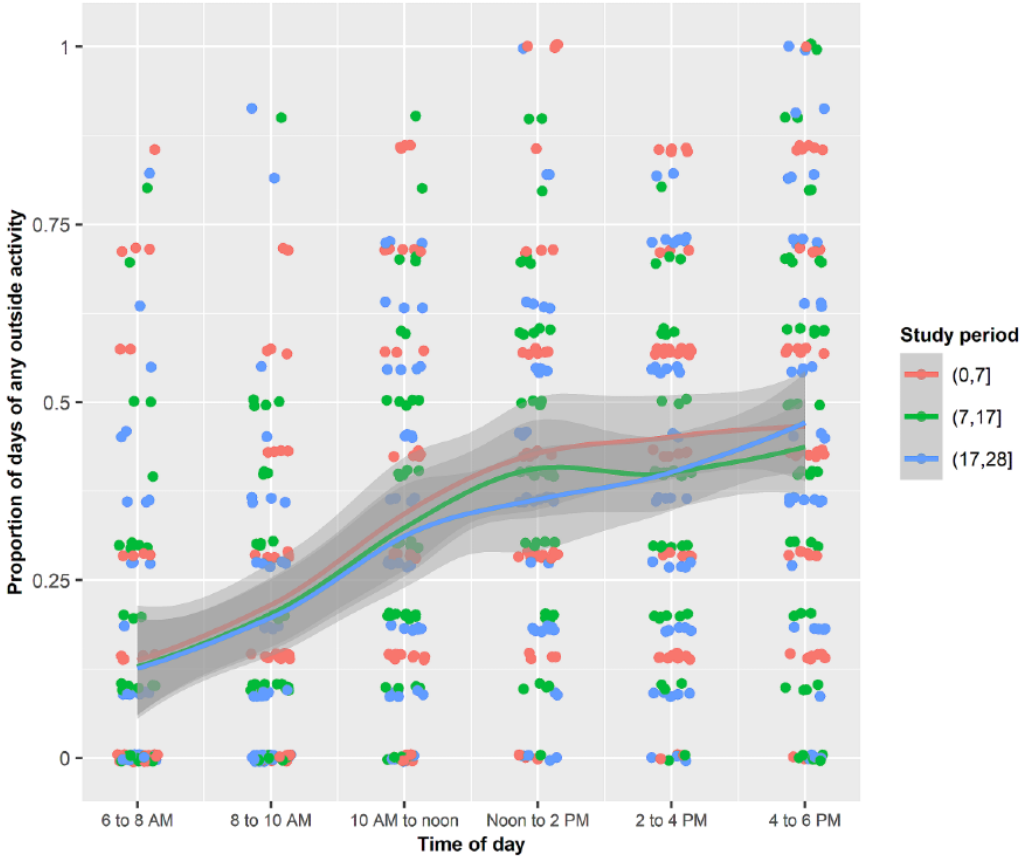


Figure 5. Mean proportions of each of the days spent outside by the sample for exercise during the 3 periods stratified by 2-hour periods.

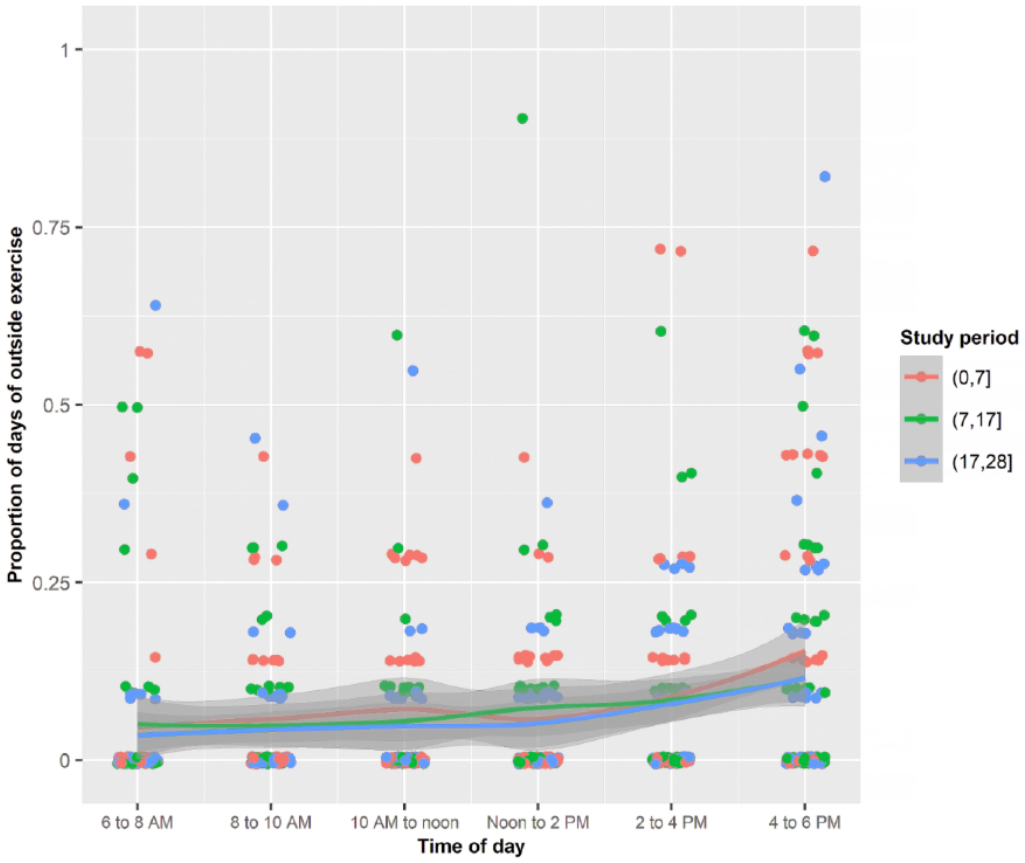


Figure 6. Mean proportions of each of the days spent outside by the sample relaxing during the 3 periods stratified by 2-hour periods.

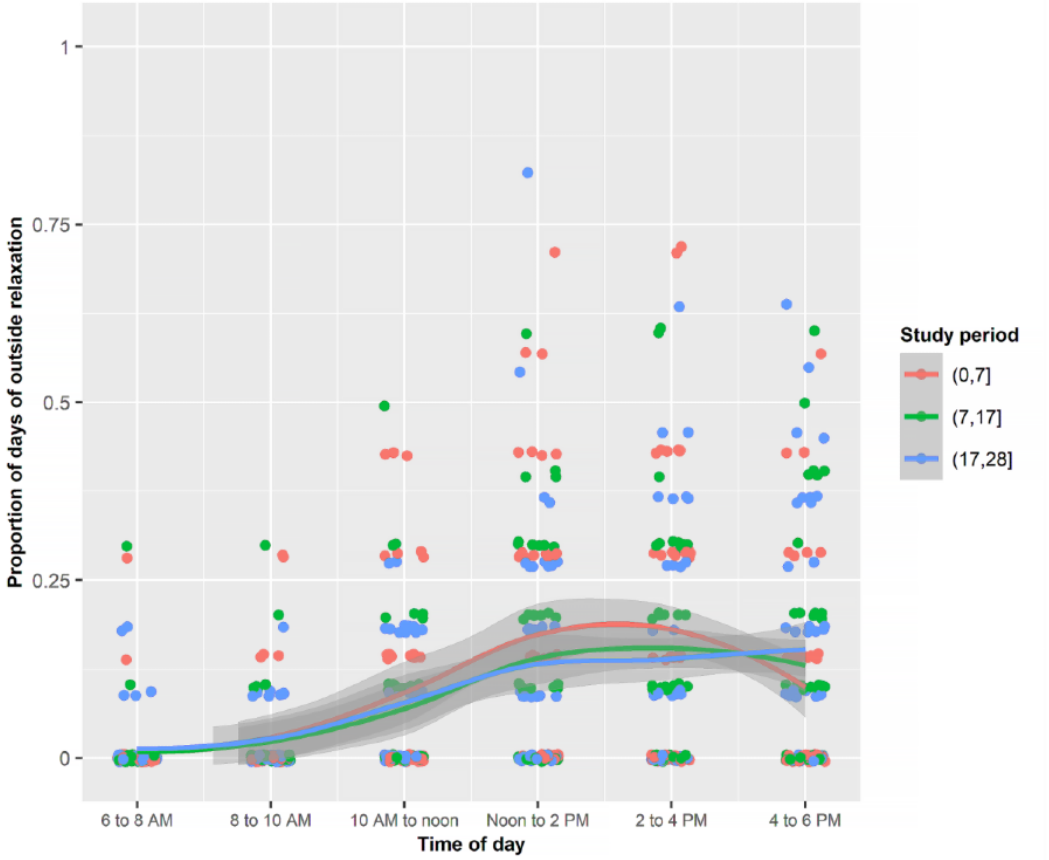


Figure 7. Mean proportions of each of the days spent outside by the sample walking during the 3 periods stratified by 2-hour periods.

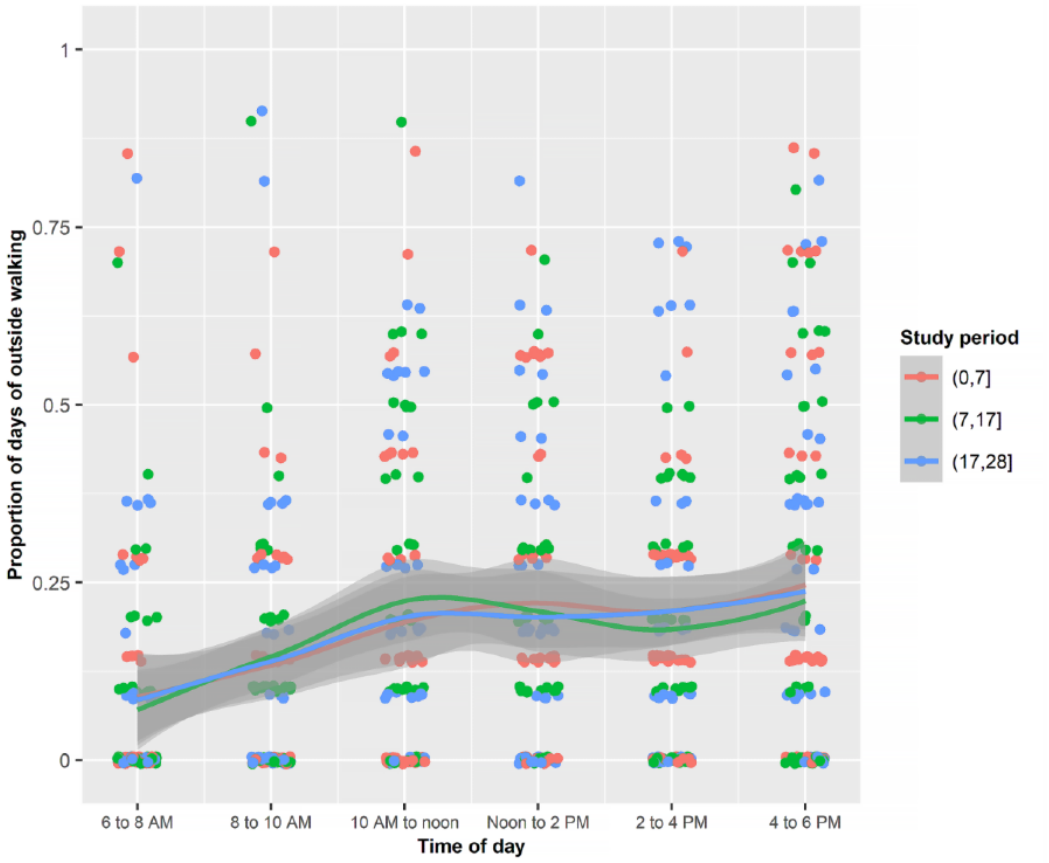
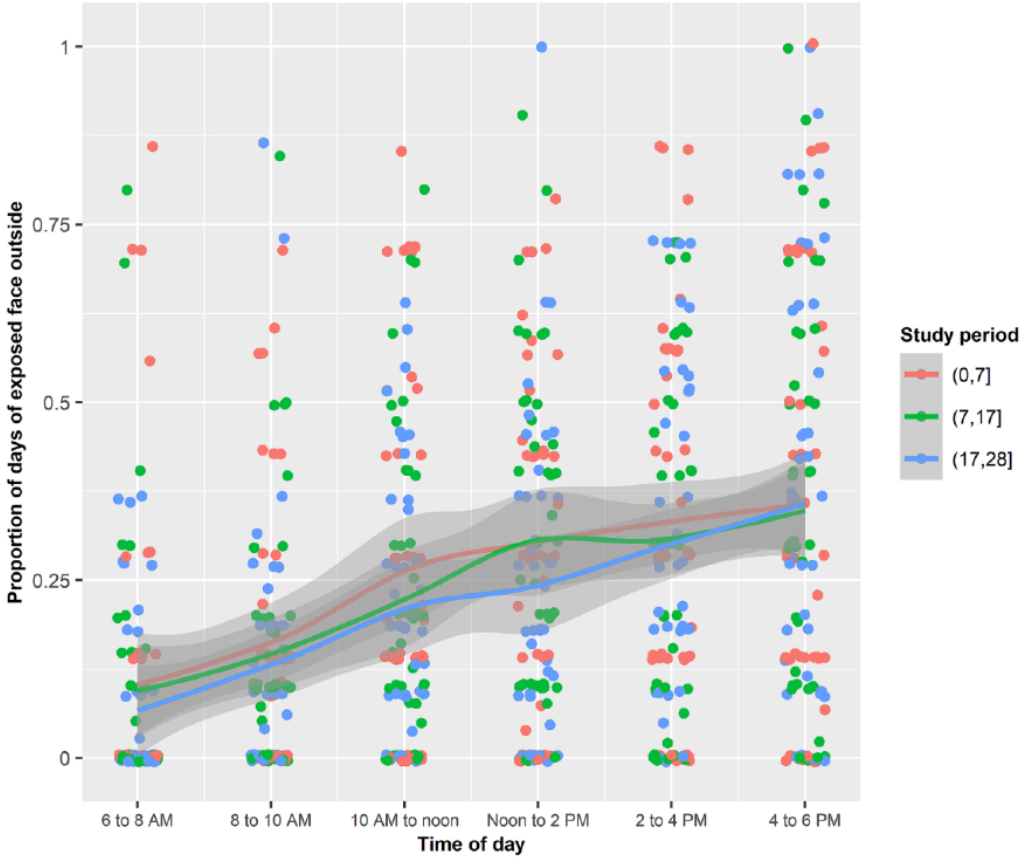


Figure 8. Mean proportions of each of the days spent outside by the sample with an unprotected face during the 3 periods stratified by 2-hour periods.



Sun Protection

Sunscreen was the principal form of sun protection used, with 73% (31/42) of participants reporting applying it each day when they went outside before noon (24/42, 58%) and in the afternoon (18/42, 42%). Most participants used one type of sunscreen (37/42, 88%) with an SPF 15-49 (23/42, 55%) or SPF 50+ (18/42, 43%). Sunscreen was mostly applied to the face (38/42, 91%) and was reapplied to the face in the afternoon by 10% (4/42) of participants. The statistically significant reduction in shoulder and chest sunburns may be attributed to wearing a T-shirt that covered the shoulders (Table 5).

Table 5. Body regions protected during the study with sunscreen and clothing.

Body regions	Sunscreen applied (n=265), n (%)	Clothing worn (n=124), n (%)
Face	242 (91)	14 (11)
Neck	99 (37)	2 (2)
Ears	63 (24)	7 (6)
Scalp	6 (2)	24 (19)
Shoulders	41 (16)	102 (82)
Back	22 (8)	106 (86)
Chest	14 (44)	88 (71)
Stomach	31 (12)	106 (86)
Arms	80 (30)	21 (17)
Hands	51 (19)	7 (6)
Buttocks	5 (2)	85 (69)
Legs	48 (18)	42 (34)
Feet	21 (8)	62 (50)

Goals Selected

The most commonly selected goal was to wear a shirt that "covers my shoulders when I am outdoors" (25/42, 60%). Other goals were selected less frequently, as follows: (1) "Plan my outdoor activities to avoid being outside from 10 AM to 2 PM" (13/42, 31%), (2) "Apply sunscreen before I go outdoors" (12/42, 29%), (3) "Wear a hat when I am outdoors" (9/42, 21%), (4) "Apply sunscreen to all of the areas of my body that may be exposed to the sun" (8/42, 19%), (5) "Be careful not to exceed the amount of UV my skin can tolerate" (8/42, 19%), and (6) "Pay attention to the strength of the sun by checking the UV Guard report 15 minutes after I go outside" (6/42, 14%).

The intention to continue to perform the sun protection behavior varied from 86% (36/42) reporting intending to "pay attention to the strength of the sun by checking the UV Guard report 15 min after I go outside," 81% (34/42) intending to apply sunscreen to all of the areas of the body that may be exposed to the sun, 81% (34/42) intending to be careful not to exceed the amount of UV the skin can tolerate, 79% (33/42) intending to wear a hat when outdoors, 71% (30/42) intending to apply sunscreen before going outdoors, 69% (29/42) intending to plan outdoor activities to avoid being outside from 10 AM to 2 PM, and 41% (17/42) intending to wear a shirt that covers the shoulders when outdoors.

Anxiety and Confidence

Anxiety about sun protection and sun exposure tended to decrease in the 3 periods from the mean score of 32.29 (SD 6.38) in period 1 to 31.69 (SD 8.84) in period 2 and to 31.95 (SD 8.41) in period 3; however, it was not statistically significant. Confidence in their ability to protect their skin from the sun and use sunscreen with reapplication, wear protective clothing, and seek shade when outdoors tended to increase in the 3 periods from the mean score of 37.33 (SD 4.72) in period 1 to 38.45 (SD 5.17) in period 2 and to 39.45 (SD 8.18) in period 3; however, this was not statistically significant.

Exit Survey Themes

Participants were surveyed at the end of the study to evaluate their experiences with the UV Guard program. Most participants (26/42, 62%) felt that they spent less time outdoors this summer in comparison with last summer. This change in time outdoors was most commonly attributed to the COVID-19 pandemic and work.

Participants expressed interest in continuing their use of the UV sensor and the UV Guard app. A total of 69% (29/42) of the participants noted that they were either extremely or moderately willing to continue using the sun protection system as part of the research, and another 26% (11/42) reported that they were either somewhat or slightly willing to do so as well. In addition, 79% (33/42) of the participants were willing to continue using the sun protection system outside of a research setting. Among this sample of 33 users willing to continue personal use, 39% (13) would use it every day.

Participants were also asked about their sun exposure and protection habits over the course of the study. A total of 48% (20/42) of the patients noted a shift in the time they went outside to periods with less-intense UV exposure. Knowledge of real-time UV exposure, which was presented in the UV Guard app, also encouraged changes in users' habits. In particular, 33% (14/42) of the participants changed the duration of the time spent outside. In addition, 38% (16/42) of the participants changed their use of clothing for sun protection or seeking shade based on their observed UV exposure. For example, some participants chose to wear hats and T-shirts when outside midday or sought shaded areas during outdoor jogging.

Discussion

Principal Findings

This proof-of-concept research demonstrated the feasibility of young adult participants using a sun protection system consisting of a personal UV-B dosimeter (UV Guard) providing real-time feedback about their personal UV exposure as graphs on their smartphone and receiving daily text messages for 28 summer days. The unique daily assessment of sunburns sustained in various regions of the body and sun protection usage on body regions made it possible to observe that sunburns of the shoulders and chest significantly decreased from the period before and after the study period, and the trend for sunburns of the arms and hands increased. Although the daily UV-B dose declined over the 3 periods, the reduction was not statistically significant. In exit surveys, 33% (14/42) of the participants reduced their cumulative daily UV-B exposure and 38% (16/42) increased their use of sun protection.

The mean daily UV-B dose during the 3 periods, which ranged from 91.96 J/m^2 to 62.73 J/m^2, exceeded the recommended daily occupational exposure limit of 30 J/m^2 within an 8-hour time frame for sensitive, unprotected skin [23]. Thus, sunburn may be expected in body areas not protected by clothing or the application of sunscreen. Defining sunburn in the recruitment materials was an important initial step in improving awareness and reporting of sunburn [10]. The reduction in facial sunburns during the study compared with those during the period preceding the intervention was attributed to improved sunscreen usage, reduced scalp sunburn to wearing a baseball cap, and reduced shoulder sunburn to wearing a T-shirt that covered the shoulders. The median age of the participants in this study (22 years) may have contributed to the lack of improvement in sunburn on the arms and hands during the study. People in the age range of 18-24 years have difficulty diminishing risk-taking behaviors, such as unprotected exposure to UV [24]. Interventions in this age group such as attempts to stop binge drinking and unsafe sex have an underwhelming track record.

Although daily tracking of sun protection behavior was repetitive, the repetition may have facilitated cognitive processing and the development of healthy habits. The daily surveys provided line drawings of types of clothing and explained the amount of sun protection provided, thus increasing knowledge and awareness of sun protection. The daily text messages initially provided knowledge about the types of sun protection and times of the day when sun protection was most necessary. Daily messages were received at the most advantageous time selected by young adults (11 AM). Working

from home during the pandemic may have skewed the selection of the time of the day to receive text messages. Young adults may have selected 11 AM as the time to receive the text as the proportion of participants going outdoors increased from 25% to 50% from 10 AM to noon. Without the constraints of the pandemic, others may prefer receiving messages earlier in the day before leaving home.

Limitations

Although this study has several strengths, response bias may have occurred, resulting in the overestimation of sun protection use. The effects of social desirability bias are expected to be minimal because of remote recruitment, provision of the intervention, and data collection. Initiation and maintenance of health promotion behaviors such as outdoor exercise and sun protection were disrupted by COVID-19, which required social-distancing policies and stay-at-home practices. Thus, the sun exposure and protection reported by the participants in this study may not be generalizable after cessation of the COVID-19 policies.

Technical limitations in the design of the UV Guard app and dosimeter may have contributed to less definitive results. As the device relies on Bluetooth communication with the UV Guard app, connectivity was hindered when the dosimeter was outside the range of the user's phone. Some participants reported that they did not carry their phones with them during outdoor exercise. In these instances, users may not have been informed with a real-time measure of sun exposure.

In addition, the sun protection system aimed to warn users as they approached their sunburn thresholds. However, participants did not receive push notifications from the UV Guard app to alert them of this risk. Rather, users were required to manually open the app to view their real-time UV exposure.

Comparison With Prior Work

Earlier efforts to reduce unprotected UV exposure among young adults focused on (1) a single point-in-time intervention such as counseling during group meetings and providing print material with appearance-focused messages to promote sun protection [25]; (2) physician counseling, including appearance-focused messages stressing the aging effects of UV on the skin and skin cancer prevention [26]; or (3) the personalization of the risk of UV exposure with a UV photograph of the participants' face showing photodamage, which resulted in increased sun protection among young adults 4-5 months and 12 months later [27]. As smartphones became ubiquitous, apps provided individuals with tailored data about their UV exposure risk based on skin sun sensitivity, including current and forecasted UVIs (Solar Cell [Klein Buendel, Inc]) [28], Healthy Texts (Cancer Australia) [29,30], and UV4.me (Rutgers Cancer Institute of New Jersey) [31]. In the 7-week interim analysis of SolarCell, there was an increase in the use of wide-brimmed hats among younger app users (24% vs 17%; P=.045), but the trend did not remain significant by the 12-week posttest analysis [32]. The approach of using a UVI based on weather predictions in a geographic region may have been impaired by people's poor overall comprehension of the UVI [33]. Studies using similar technology-based tailored text messages improved sun protection behaviors in the population studied [29,30,34].

Personal UV dosimeters and real-time UV detection communicated to users with smartphones allow interventions to be delivered by eHealth. An advantage of delivering health promotion via these mobile devices was that an already existing infrastructure was used. The Hacker et al [8] study randomized participants received either (1) the SunSmart app (Cancer Council, Victoria), (2) the UV dosimeter (Healthtronics SunSafe Pty Ltd) providing feedback set to their skin type, or (3) control with no intervention. Participants completed daily sun diaries for 4 weeks. Users received a warning when their UV-B exposure approached their sunburn threshold. The randomized controlled trial by Hacker et al [8], which gave the user their personal real-time UV dose, demonstrated a reduction in unprotected time among those using UV monitors. Our research provided an app, daily text messages, and UV dosimeter with real-time personal UV dose; therefore, relevance to the user may have been enhanced by the text messages, daily surveys, and real-time personal UV exposure provided with this high-intensity daily intervention. To our knowledge, no other trial has performed daily assessment of sunburn sustained in various regions of the body and aligned this with self-reported sun protection at 2-hour intervals. Although participants in the Hacker et al [35] study completed daily sun diaries, sun protection was recorded in less precise intervals and sunburn locations were not specified. In addition, the nature of participants' outdoor activities was not recorded in detail.

Although self-monitoring devices to assess disease treatment adherence are widespread, primary prevention behavioral change interventions are less prevalent and usually focus on changing nutrition and physical activity [36]. Evaluation of the effectiveness of physical activity and nutrition interventions has been limited by short follow-up periods and lack of patient-important primary outcomes. Although intention to change behavior and behavioral changes may be initiated during high-intensity intervention, long-term maintenance of sun protection behaviors may not be sustainable. In some regions of the United States, seasonal changes in UV intensity limit the need for sun protection in winter, thus interrupting daily reinforcement of sun protection habits. Furthermore, preventive behaviors are impacted differently because of their underlying motivations, for example, sunscreen use is influenced by social norms and attitudes [37]. Sun exposure may be related to an individual's daily routine activities, such as sitting outside to eat lunch. Our hypothesis is that as participants perform customary outdoor activities, they have sun protection habits associated with those routine activities; therefore, they will be less likely to get sunburn during routine outdoor activities. When participants sporadically engage in outdoor activities, sun protection will not be used because the habit is not established for the sporadic activity and sunburn occurs. It may not be realistic to expect sun protection and the reduction of sun exposure sustained during sporadic outdoor activities by young adults to be facilitated by mobile health with wearable sensors and mobile technologies.

Conclusions

This study demonstrated the feasibility of providing real-time UV-B exposure relative to the participants' anticipated sunburn threshold. Exit interviews and a statistically significant reduction in shoulder and chest sunburn, as well as trends in reducing sunburn of the scalp, and back indicated that the UV Guard program improved shifting outdoor activities to periods with less-intense UV exposure and wearing baseball caps and T-shirts among some young adults. Although feasibility was demonstrated, the UV Guard program needed to be optimized by having the app provide in-the-moment alarms of impending sunburn communicated by a buzzer or voice message to the user's smartphone or incorporated into the device as a vibratory, audible, or light signal. The time required to cross the sunburn threshold needs to be adjusted for the daily report of sunscreen application and clothing worn.

This study was intended to prevent sunburn and improve sun protection by providing knowledge-based daily tips and suggestions for goal setting that promote self-efficacy. Future research will explore whether young adults use the feedback provided by real-time personal UV dosimeters to stay below their daily sunburn threshold with the unintended consequence of increasing their weekly cumulative UV exposure, which is associated with increased photoaging and skin cancer. Furthermore, in the era of the COVID-19 pandemic, patterns of health behavior have been disrupted by social-distancing and sheltering-in-place policies. Future sun protection research among young adults with sun-sensitive skin will need to be postponed until the COVID-19 policies abate.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute (grant R44 CA224658) to JKR and JAR and by the National Cancer Institute Cancer of the National Institutes of Health under award number P30CA060553 to the Robert H. Lurie Comprehensive Cancer Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

There are no conflicts or competing interests for JKR, SP, ZC, EG, SYH, KK, JM, JT, JL, and MK. AB and JAR declare a relationship with Wearifi Inc and a patent on the device (US Patent Application 15/578,602 and US Patent Application 15/578,617).

Multimedia Appendix 1
The web-based consent provided to participants.
[PDF File (Adobe PDF File), 50 KB - mhealth_v9i5e25895_app1.pdf]

Multimedia Appendix 2
The schedule of measures provided to participants during the 28-day study and the web-based measures used in the study. The web-based measures are (1) baseline measures, including demographic responses; the recall of sun exposure, sunburn experienced, and sun protection used in the prior 28 days; and knowledge of sun intensity and protection; (2) measures repeated at intervals, including anxiety and confidence and daily sun habits, midpoint goal setting, and the follow-up of intentions to perform goals; and (3) exit items, including the system usability scale, questions comparing the summer experience of going outdoors in 2019 and 2020, the use of UV sensors during the study, and the willingness to continue to use the UV sensor and UV Guard app.
[PDF File (Adobe PDF File), 802 KB - mhealth_v9i5e25895_app2.pdf]

References

1. Adolescents and young adults with cancer. National Cancer Institute. URL: https://www.cancer.gov/types/aya [accessed 2020-10-05]
2. Bleyer A, Barr R. Cancer in young adults 20 to 39 years of age: overview. Semin Oncol 2009 Jun;36(3):194-206. [doi: 10.1053/j.seminoncol.2009.03.003] [Medline: 19460577]
3. Moore SC, Lee I, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern Med 2016 Jun 01;176(6):816-825 [FREE Full text] [doi: 10.1001/jamainternmed.2016.1548] [Medline: 27183032]
4. Holman DM, Berkowitz Z, Guy Jr GP, Hartman AM, Perna FM. The association between demographic and behavioral characteristics and sunburn among U.S. adults - National Health Interview Survey, 2010. Prev Med 2014 Jun;63:6-12 [FREE Full text] [doi: 10.1016/j.ypmed.2014.02.018] [Medline: 24589442]
5. Dennis LK, Vanbeek MJ, Freeman LE, Smith BJ, Dawson DV, Coughlin JA. Sunburns and risk of cutaneous melanoma: does age matter? A comprehensive meta-analysis. Ann Epidemiol 2008 Aug;18(8):614-627 [FREE Full text] [doi: 10.1016/j.annepidem.2008.04.006] [Medline: 18652979]
6. Holman DM, Ding H, Guy GP, Watson M, Hartman AM, Perna FM. Prevalence of sun protection use and sunburn and association of demographic and behaviorial characteristics with sunburn among US adults. JAMA Dermatol 2018 May 01;154(5):561-568 [FREE Full text] [doi: 10.1001/jamadermatol.2018.0028] [Medline: 29541756]

7. Heckman CJ, Darlow SD, Ritterband LM, Handorf EA, Manne SL. Efficacy of an intervention to alter skin cancer risk behaviors in young adults. Am J Prev Med 2016 Jul;51(1):1-11 [FREE Full text] [doi: 10.1016/j.amepre.2015.11.008] [Medline: 26810358]
8. Hacker E, Horsham C, Vagenas D, Jones L, Lowe J, Janda M. A mobile technology intervention with ultraviolet radiation dosimeters and smartphone apps for skin cancer prevention in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Nov 28;6(11):e199 [FREE Full text] [doi: 10.2196/mhealth.9854] [Medline: 30487115]
9. Robinson JK, Friedewald JJ, Desai A, Gordon EJ. A randomized controlled trial of a mobile medical app for kidney transplant recipients. Transplant Direct 2016;2(1):e51. [doi: 10.1097/TXD.0000000000000561]
10. Robinson JK, Durst DA, Gray E, Kwasny M. Protection-adjusted UV dose estimated for body areas: daily self-reported sun protection modification of wearable UV sensor dose. Photodermatol Photoimmunol Photomed 2020 Sep 28;36(5):357-364. [doi: 10.1111/phpp.12557] [Medline: 32189399]
11. Xu S, Stump TK, Jain J, Alshurafa N, Robinson JK. Variation in daily ultraviolet radiation exposure and sun protection behaviours of melanoma survivors: an observational single-arm pilot study with a wearable sensor. Br J Dermatol 2019 Feb 16;180(2):413-414 [FREE Full text] [doi: 10.1111/bjd.17196] [Medline: 30207589]
12. 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]
13. Kwon K, Heo SY, Yoo I, Banks A, Chan M, Lee JY, et al. Miniaturized, light-adaptive, wireless dosimeters autonomously monitor exposure to electromagnetic radiation. Sci Adv 2019 Dec 13;5(12):aay2462 [FREE Full text] [doi: 10.1126/sciadv.aay2462] [Medline: 31853499]
14. Heo SY, Kim J, Gutruf P, Banks A, Wei P, Pielak R, et al. Wireless, battery-free, flexible, miniaturized dosimeters monitor exposure to solar radiation and to light for phototherapy. Sci Transl Med 2018 Dec 05;10(470):eaau1643 [FREE Full text] [doi: 10.1126/scitranslmed.aau1643] [Medline: 30518611]
15. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Hoboken, New Jersey, United States: Prentice Hall; 1985:1-640.
16. Robinson JK, Durst DA, Gray E, Kwasny M, Heo SY, Banks A, et al. Sun exposure reduction by melanoma survivors with wearable sensor providing real-time UV exposure and daily text messages with structured goal setting. Arch Dermatol Res 2020 Nov 13. [doi: 10.1007/s00403-020-02163-1] [Medline: 33185716]
17. Diffey B. Sun Protection: A Risk Management Approach. Bristol, UK: Institute of Physics Publishing; 2017:317-318.
18. Ouinn A, Diffey B, Craig P, Farr P. Definition of the minimal erythemal dose used for diagnostic phototesting. Br J Dermatol 1994 Jul;131(s44):56. [doi: 10.1111/j.1365-2133.1994.tb16151.x]
19. Robinson JK, Guevara Y, Gaber R, Clayman ML, Kwasny MJ, Friedewald JJ, et al. Efficacy of a sun protection workbook for kidney transplant recipients: a randomized controlled trial of a culturally sensitive educational intervention. Am J Transplant 2014 Dec;14(12):2821-2829 [FREE Full text] [doi: 10.1111/ajt.12932] [Medline: 25395386]
20. The R Project for statistical computing. The R Foundation. URL: https://www.R-project.org/ [accessed 2021-04-14]
21. Fioletov V, Kerr JB, Fergusson A. The UV index: definition, distribution and factors affecting it. Can J Public Health 2010 Jul 1;101(4):5-9. [doi: 10.1007/bf03405303]
22. Diffey BL. Time and place as modifiers of personal UV exposure. Int J Environ Res Public Health 2018 May 30;15(6):1112 [FREE Full text] [doi: 10.3390/ijerph15061112] [Medline: 29848972]
23. International Commission on Non-Ionizing Radiation Protection. ICNIRP statement--Protection of workers against ultraviolet radiation. Health Phys 2010 Jul;99(1):66-87. [doi: 10.1097/HP.0b013e3181d85908] [Medline: 20539126]
24. Steinberg LD. Age of Opportunity: Lessons From the New Science of Adolescence. Boston MA: Houghton Miffline Harcourt; 2015.
25. Hillhouse J, Turrisi R, Stapleton J, Robinson J. A randomized controlled trial of an appearance-focused intervention to prevent skin cancer. Cancer 2008 Dec 01;113(11):3257-3266 [FREE Full text] [doi: 10.1002/cncr.23922] [Medline: 18937268]
26. US Preventive Services Task Force, Grossman DC, Curry SJ, Owens DK, Barry MJ, Caughey AB, et al. Behavioral counseling to prevent skin cancer: US preventive services task force recommendation statement. J Am Med Assoc 2018 Mar 20;319(11):1134-1142. [doi: 10.1001/jama.2018.1623] [Medline: 29558558]
27. Mahler HI, Kulik JA, Gerrard M, Gibbons FX. Long-term effects of appearance-based interventions on sun protection behaviors. Health Psychol 2007 May;26(3):350-360. [doi: 10.1037/0278-6133.26.3.350] [Medline: 17500622]
28. Buller DB, Berwick M, Lantz K, Buller MK, Shane J, Kane I, et al. Smartphone mobile application delivering personalized, real-time sun protection advice: a randomized clinical trial. JAMA Dermatol 2015 May 01;151(5):497-504 [FREE Full text] [doi: 10.1001/jamadermatol.2014.3889] [Medline: 25629710]
29. Janda M, Youl P, Marshall AL, Soyer HP, Baade P. The HealthyTexts study: a randomized controlled trial to improve skin cancer prevention behaviors among young people. Contemp Clin Trials 2013 May;35(1):159-167. [doi: 10.1016/j.cct.2013.03.009] [Medline: 23557730]

30. Youl PH, Soyer HP, Baade PD, Marshall AL, Finch L, Janda M. Can skin cancer prevention and early detection be improved via mobile phone text messaging? A randomised, attention control trial. Prev Med 2015 Feb;71:50-56. [doi: 10.1016/j.ypmed.2014.12.009] [Medline: 25524612]
31. Heckman CJ, Handorf EA, Darlow SD, Ritterband LM, Manne SL. An online skin cancer risk-reduction intervention for young adults: mechanisms of effects. Health Psychol 2017 Mar;36(3):215-225 [FREE Full text] [doi: 10.1037/hea0000420] [Medline: 27819460]
32. Buller DB, Berwick M, Lantz K, Buller MK, Shane J, Kane I, et al. Evaluation of immediate and 12-week effects of a smartphone sun-safety mobile application: a randomized clinical trial. JAMA Dermatol 2015 May 01;151(5):505-512 [FREE Full text] [doi: 10.1001/jamadermatol.2014.3894] [Medline: 25629819]
33. Heckman CJ, Liang K, Riley M. Awareness, understanding, use, and impact of the UV index: a systematic review of over two decades of international research. Prev Med 2019 Jun;123:71-83 [FREE Full text] [doi: 10.1016/j.ypmed.2019.03.004] [Medline: 30844501]
34. Ho BK, Reidy K, Huerta I, Dilley K, Crawford S, Hultgren BA, et al. Effectiveness of a multicomponent sun protection program for young children: a randomized clinical trial. JAMA Pediatr 2016 Apr 01;170(4):334-342 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.4373] [Medline: 26857829]
35. Hacker E, Horsham C, Allen M, Nathan A, Lowe J, Janda M. Capturing ultraviolet radiation exposure and physical activity: feasibility study and comparison between self-reports, mobile apps, dosimeters, and accelerometers. JMIR Res Protoc 2018 Apr 17;7(4):e102 [FREE Full text] [doi: 10.2196/resprot.9695] [Medline: 29666044]
36. Oikonomidi T, Vivot A, Tran V, Riveros C, Robin E, Ravaud P. A methodologic systematic review of mobile health behavior change randomized trials. Am J Prev Med 2019 Dec;57(6):836-843. [doi: 10.1016/j.amepre.2019.07.008] [Medline: 31753266]
37. Kasparian NA, McLoone JK, Meiser B. Skin cancer-related prevention and screening behaviors: a review of the literature. J Behav Med 2009 Oct;32(5):406-428. [doi: 10.1007/s10865-009-9219-2] [Medline: 19521760]

Abbreviations

MED: minimal erythema dose
OR: odds ratio
REDCap: Research Electronic Data Capture
SPF: sun protection factor
UVI: ultraviolet index

Edited by L Buis; submitted 19.11.20; peer-reviewed by A Cust, C Heckman; comments to author 31.12.20; revised version received 22.01.21; accepted 05.04.21; published 06.05.21.

Please cite as:
Robinson JK, Patel S, Heo SY, Gray E, Lim J, Kwon K, Christiansen Z, Model J, Trueb J, Banks A, Kwasny M, Rogers JA
Real-Time UV Measurement With a Sun Protection System for Warning Young Adults About Sunburn: Prospective Cohort Study
JMIR Mhealth Uhealth 2021;9(5):e25895
URL: https://mhealth.jmir.org/2021/5/e25895
doi:10.2196/25895
PMID:33955844

Original Paper

Commercially Available Apps to Support Healthy Family Meals: User Testing of App Utility, Acceptability, and Engagement

Chelsea E Mauch[1,2], PhD; Rachel A Laws[2,3], PhD; Ivanka Prichard[1,4], PhD; Anthony J Maeder[1,5], PhD; Thomas P Wycherley[6], PhD; Rebecca K Golley[1,2], PhD

[1]Caring Futures Institute, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[2]Early Prevention of Obesity in Childhood Centre of Research Excellence, Sydney, Australia

[3]School of Exercise and Nutrition Science, Institute for Physical Activity and Nutrition, Deakin University, Melbourne, Australia

[4]Sport, Health, Activity, Performance and Exercise Research Centre, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[5]Flinders Digital Health Research Centre, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[6]Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

Corresponding Author:
Chelsea E Mauch, PhD
Caring Futures Institute
College of Nursing and Health Sciences
Flinders University
GPO Box 2100
Adelaide, 5001
Australia
Phone: 61 8 83038988
Email: chelsea.mauch@csiro.au

Abstract

Background: Parents juggling caregiving and paid employment encounter a range of barriers in providing healthy food to their families. Mobile apps have the potential to help parents in planning, purchasing, and preparing healthy family food. The utility and acceptability of apps for supporting parents are unknown. User perspectives of existing technology, such as commercially available apps, can guide the development of evidence-based apps in the future.

Objective: This study aims to determine the feasibility of existing commercially available apps for supporting the healthy food provision practices of working parents.

Methods: Working parents (N=133) were recruited via the web and completed a 10-item Capability, Opportunity, Motivation, and Behavior (COM-B) self-evaluation survey assessing their needs in relation to the provision of healthy family meals. A total of 5 apps were selected for testing, including a meal planning app, recipe app, recipe manager app, family organizer app, and barcode scanning app. Survey items were mapped to app features, with a subsample of parents (67/133, 50.4%) allocated 2 apps each to trial simultaneously over 4 weeks. A semistructured interview exploring app utility and acceptability and a web-based survey, including the System Usability Scale and the user version of the Mobile App Rating Scale, followed app testing. The interview data were analyzed using a theoretical thematic approach.

Results: Survey participants (N=133; mean age 34 years, SD 4 years) were mainly mothers (130/133, 97.7%) and partnered (122/133, 91.7%). Participants identified a need for healthy recipes (109/133, 82% agreed or strongly agreed) and time for food provision processes (107/133, 80.5%). Engagement quality was the lowest rated domain of the user version of the Mobile App Rating Scale across all 5 apps (mean score per app ranging from 3.0 to 3.7 out of a maximum of 5). The family organizer, requiring a high level of user input, was rated the lowest for usability (median 48, IQR 34-73). In the interviews, participants weighed the benefits of the apps (ie, time saving) against the effort involved in using them in determining their acceptability. Organization was a subtheme emerging from interviews, associated with the use of meal planners and shopping lists. Meal planners and shopping lists were used in time, while behavior was occurring.

Conclusions: Meal planning apps and features promoting organization present feasible, time-saving solutions to support healthy food provision practices. Attention must be paid to enhancing app automation and integration, as well as recipe and nutrition content, to ensure that apps do not add to the time burden of food provision and are supportive of healthy food provision behavior in time.

(JMIR Mhealth Uhealth 2021;9(5):e22990) doi:10.2196/22990

KEYWORDS

diet; nutrition; family; mobile applications; behavior modification

Introduction

Background

Suboptimal dietary intake is a major public health concern because of its role in the development of noncommunicable diseases (NCDs) [1]. In 2016, NCDs were responsible for 70% of deaths worldwide [2]. Key dietary risk factors for NCDs include inadequate intake of vegetables, fruit, and wholegrains and excessive intake of energy-dense, nutrient-poor foods, also termed discretionary choices [1]. A total of 90% of Australian adults do not meet the recommended daily serves of vegetables, and more than a third of daily energy intake is from discretionary choices [3]. Similar trends have been observed internationally [4,5]. Poor dietary patterns start young and persist over time. Australian children's diet quality mirrors adult patterns by 4-8 years of age [3,6]. Supporting parents to provide healthy food to themselves and their families will improve population diet quality.

There has been a trend toward greater female workforce participation in modern households [7,8]. In Australia in 2019, 70% of mothers in dual-parent households were working, whereas 60% of mothers in single-parent households were working [7]. Parents juggling caregiving and paid employment experience a range of barriers in providing healthy food to their families. The Capability, Opportunity, Motivation, and Behavior (COM-B) system [9] describes 3 key conditions that interact to enable a behavior to occur: capability, opportunity, and motivation. Parent-focused nutrition interventions to date have tended to target capability (eg, knowledge and skills) and motivation (eg, confidence in supporting child health) [10]. However, opportunity-related enablers, such as adequate time for food provision, are important and promote resilience against the broader unhealthy food environment [11,12]. Therefore, it is important to consider a range of enablers relevant to the planning, purchasing, and preparation of food in the development of future nutrition interventions.

The time- and staff-intensive nature of traditional face-to-face interventions make them impractical in a resource-scarce health promotion environment. Mobile apps offer advantages over face-to-face interventions, such as the delivery of interventions in everyday situations [13]. A review identified 51 commercially available apps that addressed the planning, purchasing, and preparation of food [14]. The review found that meal planning, family organizer, and recipe manager apps incorporated features promoting organization that could address potential barriers to healthy meal provision, such as time scarcity and cognitive load [14]. However, app content generally mapped to relatively few behavior change techniques and was not targeted toward healthy eating in a family context [14].

Objectives

The next step in understanding the behavioral potential of these types of apps and features in a family food provision context is to gain insights from target users. User perspectives can inform the design of evidence-based apps that are informed by user context and needs [15-17]. This study sought user perspectives on commercially available apps to inform future app development or refinement [18]. This study aims to determine the feasibility of existing commercially available apps and app features for supporting healthy food provision practices in working parents by exploring the following:

1. The utility of apps and app features to support planning, purchasing, and preparation of food
2. The acceptability of apps and app features in terms of quality, usability, functionality, and engagement.

Methods

Study Design

This feasibility study was conducted between February and June 2019 using a mixed methods design. Participants completed a baseline survey, with a subsample undertaking a 4-week app testing period, followed by another survey and semistructured interview. A total of 5 apps were selected for testing based on a previous review of commercially available apps [14]. Selected apps represented the key content and features of interest identified in the previous review (Multimedia Appendix 1 [14]). They rated well for quality compared with similar apps, were available in a free or freemium format, and were available on Apple and Android operating systems. Only one of the apps tested has published research available regarding its development [19].

Study Sample and Recruitment

Eligibility criteria included being a single or partnered parent in paid employment, with themselves or their partner having returned to work from a period of parental leave in the last 6 months. Other eligibility criteria included being based in Australia and the main food gatekeeper of the household. Individuals who did not own an Apple or Android mobile device with internet access or whose partner was not in paid employment were excluded.

Recruitment was conducted via Facebook and flyers posted around a university campus and in childcare centers. These recruitment channels have been used successfully in previous research [20-22]. Baseline survey completion constituted consent for the survey only. Participants provided contact details at the end of the survey to indicate their interest in app testing. Consent for app testing was by return email, with reminders sent to nonresponders until recruitment and app allocation goals were met. A target sample size of 50 was set for app testing, with at least 10 participants testing each app. This was comparable with similar feasibility and pilot app testing studies [22-24]. Ethics approval was provided by the Flinders University Social and Behavioral Research Ethics Committee (approval no. 8211).

App Testing

A total of 10 baseline questions modeled on the COM-B self-evaluation survey [25] exploring the perceived enablers of healthy food provision were mapped to the apps for testing (Multimedia Appendix 1). Participants were assigned 2 of the 5 mobile apps for testing, based on their responses to the baseline questions. This allowed the allocation of apps based on need. The allocation of 2 apps allowed participants to envisage how complementary content and features could be combined. Consenting participants were contacted via telephone or email for app allocation and setup and emailed a checklist of tasks to prompt use of a range of app features (eg, viewing a recipe, creating a new meal plan, and setting a day and time to receive meal planning reminders). They were encouraged to use both the apps as little or as much as they wished during the following 4 weeks.

Follow-up

At the completion of the 4-week app testing period, participants were emailed a link to the follow-up survey, with 1 to 3 reminder emails sent to nonresponders. Following receipt of the follow-up survey, participants were contacted by telephone to conduct a semistructured interview (until data saturation was reached). Participants involved in app testing were provided with a meal kit or grocery voucher to the value of Aus $85 (US $66) in compensation for their time.

Data Collection

Baseline Survey

Demographic survey items included parents' age, sex, highest level of education, relationship status, household income, and work hours; partner's work hours (if applicable); and the number and age of children in the household. Diet quality measures were adapted from the validated Short Food Survey [26] and included questions relating to fruits and vegetables (2 items) and discretionary choice intake (10 items). For discretionary choice items, frequency of consumption (ie, daily, weekly, and monthly) was followed by a question regarding the number of times it was usually consumed (eg, twice or thrice). Discretionary choice items included sugar-sweetened beverages, takeaway foods, fried potatoes, savory snacks, savory pastries, sweet baked goods, snack bars, confectionaries, and frozen desserts.

The 10 COM-B self-evaluation items (Multimedia Appendix 2) were rated on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7) [25]. Scores from the 2 items mapping to each app were summed, with participants allocated the 2 apps receiving the highest aggregate score. Where an app was not allocated to at least 10 participants, some participants were allocated these apps despite a lower COM-B score, to ensure that adequate data were collected for each app.

Follow-up Survey

Frequency of app use was measured in the follow-up survey for each of the 4 weeks using a 4-point response scale (ie, *didn't use the app*, *once*, *2-4 times*, and *5 or more times*). The duration of app use was measured using a 3-point response scale (ie, *less than 1 min*, *1-5 min*, and *more than 5 min*). The System Usability Scale (SUS), a brief scale of 10 statements covering the complexity or ease of use of apps, was used to assess usability [27,28]. Participants indicated their agreement on a 5-point scale ranging from strongly disagree (1) to strongly agree (5). User-perceived app quality was measured using the user version of the Mobile App Rating Scale (uMARS) [29]. The 16-items comprising the engagement, functionality, esthetics, and subjective quality subscales were included in this study [29]. The information quality subscale was replaced with an item regarding app credibility, as the included apps contained minimal information, and credibility has been shown to be important to app engagement [21].

Semistructured Interviews

The Consolidated Criteria for Reporting Qualitative Research checklist guided the presentation of qualitative methods and findings [30]. Semistructured interviews were conducted by a female research assistant with qualitative research experience and a research focus on family meals. The research assistant had no previous contact with participants. Three of the interviews were conducted by CEM, including a pilot interview and 2 final interviews. Interviews took between 30 and 60 minutes and were audio recorded with the participants' permission, using a speaker phone and audio recorder. Interviews were transcribed verbatim by an independent company.

Interview questions addressed the acceptability of app features and content, user engagement with the apps, the usefulness of the apps in addressing food provision, improvements required, and general suggestions for future app development. Questions were repeated for each of the 2 apps. The questions were tested with a research assistant and piloted with one participant. As interviews were conducted, CEM listened to the audio and discussed progress with the research assistant. Once data saturation for an app or app combination was reached, determined by no new information emerging, participants testing those apps were only asked to complete the follow-up survey. Single parents and parents of a lower income were prioritized for interviews to represent as diverse a sample as possible.

Data Analysis

Quantitative Data

Quantitative data were analyzed using SPSS version 22 (IBM Corporation). Parental work hours were converted from continuous variables into groups (ie, part time=1 to <35 hours per week and full time=35+ hours) and combined to describe the family work schedule. Discretionary items were summed as total serves per day using age-and sex-specific adjustment factors [31]. Demographic data, diet quality, and COM-B self-evaluation items were presented descriptively (eg, n [%], mean [SD], and median [IQR] for COM-B items due to a positive skew in the data).

Follow-up data regarding self-reported frequency and duration of app use were calculated for each app and presented descriptively as n (%). SUS scores were converted to a score out of 100 [27], with the median (IQR) score of the sample presented, due to a positive skew in the data. A median score below 50 was indicative of poor app usability, 50 to 70 was

marginal, more than 70 was passable, and more than 90 was superior [28]. Scores for uMARS items were summed and averaged for each subscale and across all items for the overall uMARS score.

Qualitative Data

Transcriptions were coded using NVivo (QSR International) and analyzed using a theoretical thematic approach [32]. Coding took an inductive approach, with interview data initially sorted into groups based on the study objectives, interview questions, and app characteristics. Coding was conducted by CEM, who organized the data into major and minor themes and generated an initial conceptual model [32]. A meeting between coauthors was undertaken to discuss and refine themes, after which the final conceptual model with links back to quantitative data was ascertained. When presenting the interview results, names were changed to preserve anonymity.

Results

Sample Characteristics

Figure 1 describes participant flow throughout the study. Participants completing the baseline survey (N=133) were mostly partnered (122/133, 91.7%) and females (130/133, 97.7%). Two-thirds (90/133, 67.7%) of households included one full-time working parent and one part-time working parent. In most households (106/133, 79.7%), the age of the youngest child was less than 2 years, and 60.2% (80/133) included more than one child. Only 6.8% (9/133) of participants met the Australian guidelines for vegetable intake, whereas most (107/133, 80.5%) met the guidelines for fruit intake. Participants reported consuming 3.0 (SD 2.1) discretionary choice serves per day, excluding alcohol (Table 1). Compared with the baseline sample, participants who completed interviews were more likely to have a university degree (28/36, 78% vs 83/133, 62.4%), be unpartnered (5/36, 14% vs 11/133, 8.3%), and have an income below Aus $70,000 (US $54,221) per annum (9/36, 25% vs 25/133, 18.7%; Table 1).

Table 2 presents the results of the COM-B self-evaluation items used to allocate apps. More than three-quarters of participants suggested a need for healthy recipes and meal ideas (109/133, 82% agreed or strongly agreed; median 6, IQR 6-7) and time to plan, buy, and prepare healthy meals (107/133, 80.5%; median 6, IQR 6-7). Almost two-thirds suggested a need for a better way of planning and recording meals and groceries (87/133, 65.4%; median 6, IQR 5-6), whereas food selection and cooking skills were not high priorities for this sample.

Figure 1. Flow of participants through the study.

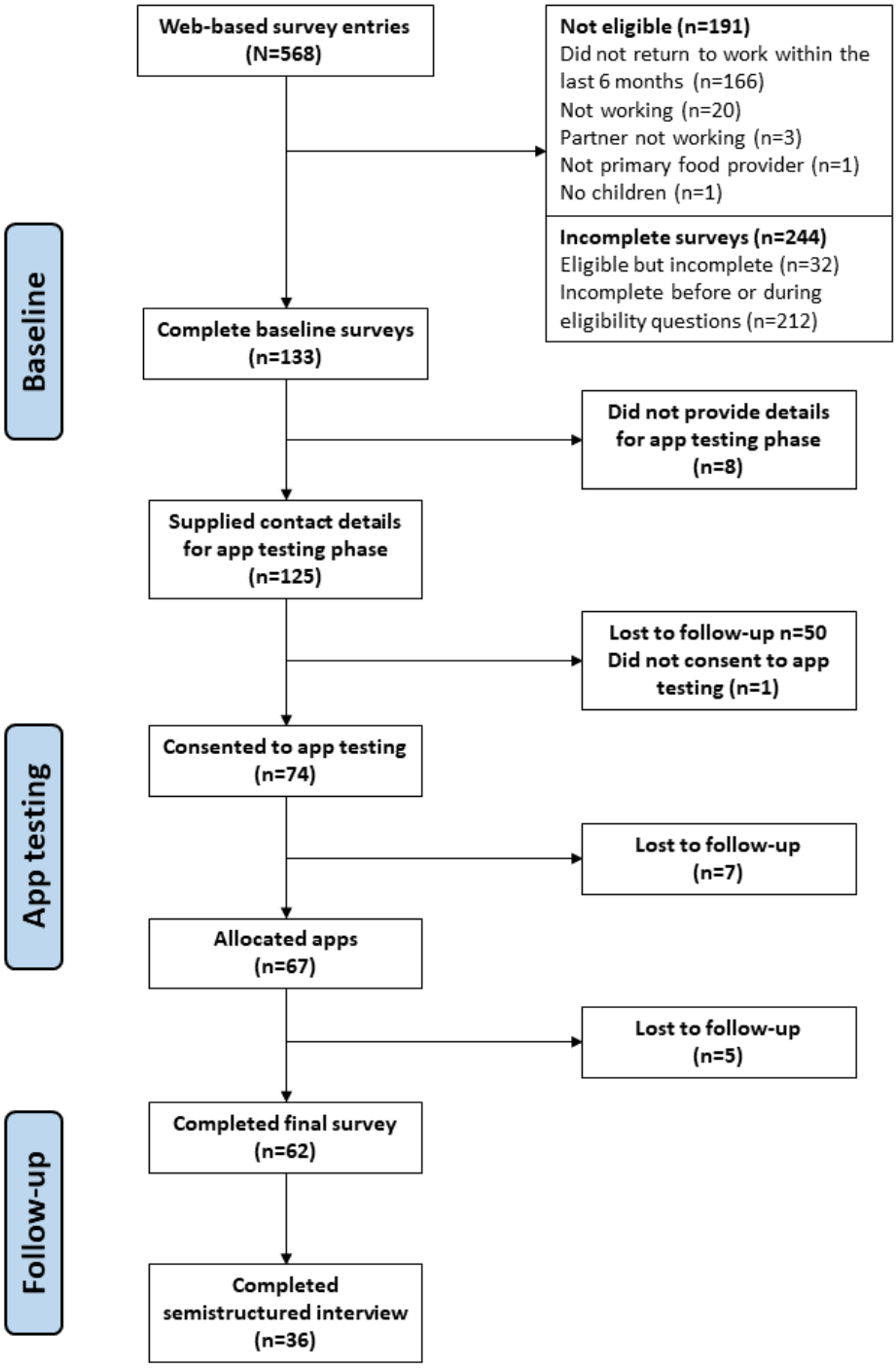


Table 1. Demographic characteristics of the survey sample at baseline and the sample included in the qualitative analysis.

Characteristics	Baseline survey data sample[a]	Interview subsample[b]
Age (years) mean (SD)	33.8 (4.3)	33.6 (4.3)
Sex, n (%)		
Female	130 (97.7)	35 (97.2)
Male	3 (2.3)	1 (2.8)
Highest level of education, n (%)		
University	83 (62.4)	28 (77.8)
No university	50 (37.6)	8 (22.2)
Relationship status, n (%)		
Partnered	122 (91.7)	31 (86.1)
Single	11 (8.3)	5 (13.9)
Number of children, n (%)		
One	50 (38.5)	13 (37.1)
More than one	80 (61.5)	22 (62.9)
Age of youngest child (years), n (%)		
Less than 2	106 (79.7)	29 (80.6)
2-4	21 (15.8)	5 (13.9)
5-12	6 (4.5)	2 (5.6)
Household income (gross per annum), n (%)		
Less than Aus $70,000 (US $54,221)	25 (18.8)	9 (25.0)
Aus $70,000 (US $54,221) or more	92 (69.2)	24 (66.7)
Prefer not to say	16 (12.0)	3 (8.3)
Family work schedule, n (%)		
Both part time	9 (6.8)	1 (2.8)
Part time and full time	90 (67.7)	26 (72.2)
Both full time	23 (17.3)	4 (11.1)
Single working parent	11 (8.3)	5 (13.9)
Vegetable intake (serves per day), n (%)		
1 or less	26 (19.5)	8 (22.2)
2-4	98 (73.7)	26 (72.2)
5 or more	9 (6.8)	2 (5.6)
Fruit intake (serves per day), n (%)		
Do not eat fruit	4 (3.0)	1 (2.8)
1 or less	69 (51.9)	17 (47.2)
2 or more	60 (45.1)	18 (50.0)
Discretionary intake (serves)[c], mean (SD)	3.0 (2.1)	3.0 (2.2)

[a]Samples range from 118 to 133 due to missing data.

[b]Subsamples range from 33 to 36 due to missing data.

[c]Excluding alcohol.

Table 2. COM-B self-evaluation item mean (SD) scores and proportions of the sample responding agreed or strongly agreed (N=133).

COM-B[a] domain and item	Item score[b], median (IQR)	Agreed or strongly agreed[c], n (%)
Capability		
Have better food preparation or cooking skills	5 (3-6)	44 (33.1)
Learn how to choose healthy food at the supermarket	5 (2-5)	30 (22.6)
Learn how to plan healthy meals	6 (5-6)	72 (54.1)
Opportunity		
Have more time to plan, buy, and prepare healthy meals	6 (6-7)	107 (80.5)
Have more healthy recipes and meal ideas	6 (6-7)	109 (82.0)
Have guidance in choosing healthy food and meals	5 (4-6)	38 (28.6)
Have a better way of planning and recording meals and groceries for the coming week	6 (5-6)	87 (65.4)
Have more support or help from my partner or family	5 (4-6)	43 (32.3)
Have more reminders to plan, shop or cook	5 (4-6)	43 (32.3)
Motivation		
Have clear goals or plans toward preparing healthy meals	6 (5-6)	76 (57.1)

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

[b]1=strongly disagree to 7=strongly agree.

[c]Score of 6 or 7 (agreed or strongly agreed).

Follow-up and Interview Data

Of the 67 participants who were allocated apps, 62 (93%) completed the follow-up survey and 36 (54%) participants completed the interviews (Figure 1). Among those completing interviews, 9 different combinations of apps were allocated. The most common sets of apps allocated were the recipe manager and family organizer (n=7) or meal planning app (n=6) and the barcode scanning and recipe app (n=6). Figure 2 demonstrates the conceptual model of major and minor themes emerging from the semistructured interviews, and Table 3 provides examples of quotes relating to each subtheme.

Figure 2. Conceptual diagram of major themes (blue) and minor themes (yellow) and how these may relate to the ongoing use or disengagement with the apps.

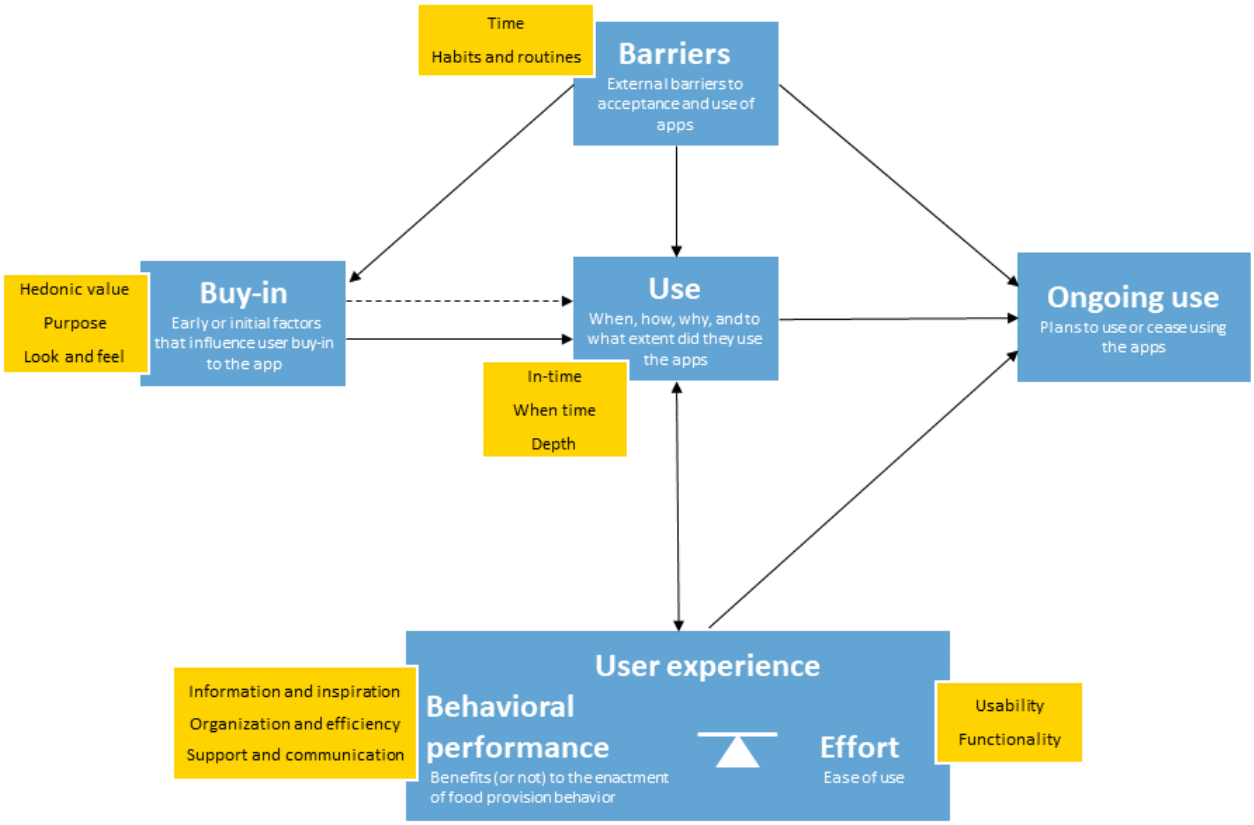


Table 3. Participants' perspectives of app utility and acceptability and their engagement with apps: major and minor themes and illustrative quotes.

Major and minor themes	Illustrative quotes
Buy-in	
Purpose	• "(...) for probably what I was looking for which was meal planning, [meal planning app] was more appropriate." (Mia, working 1 to <21 hours per week) • "...I think, it had a big overarching purpose but lots of, like, little purposes in there that just, kind of meant that you had to wade through more stuff to figure out what you wanted to use it for." (Jo, working 21 to <35 hours per week)
Hedonic value	• "...I hadn't even, um, really thought about the fact that there were apps out there to support with meal prep and healthy eating and all of that..." (Harper, working 21 to <35 hours per week) • "...I didn't want to use the app. (...) I wasn't excited by it." (Sophie, working 21 to <35 hours per week)
Look and feel	• "...you'd like trust and you feel comfort in knowing that, you know, it just feels like a team of people has worked behind it..." (Tiffany, working 21 to <35 hours per week)
Use	
In-time	• "So we would (...) decide on the meals (...) then I would go grocery shopping (...) then I would put it away until I needed to cook every night." (Blair, working 21 to <35 hours per week)
When-time	• "...when the kids were sort of asleep (...) just looked up some recipes and that type of thing so—'cause I just had some time to actually do it." (Dianne, working 1 to <21 hours per week)
Depth	• "I don't think I was even aware of it. (...) I don't think I even found that function." (Bianca, working 21 to <35 hours per week)
Barriers	
Time	• "I think, real or just perceived, I think, that's a, um, a time issue, I feel, like, (...) there's other things I should be doing..." (Cora, working 21 to <35 hours per week)
Habits and routines	• "...the app can be as brilliant as it is but if I'm not going to actually actively go out of my way to build that habit, (...) it's only as good as I'm going to make it." (Sophie, working 21 to <35 hours per week) • "...I actually ended up changing jobs, like, right in the middle of, um, trialing the app. (...) I tried to settle back to just doing what was easy..." (Harper, working 21 to <35 hours per week)
User experience — behavioral performance	
Information and instructions	• "...if I was, you know, tired or whatever, I might just turn to a freezer meal (...) whereas this kind of made me think, like making it from scratch and like, still find easy ways to get, you know, vegetables into my kids" (Kathryn, working 1 to <21 hours per week) • "...aligned to the Australian (...) food guidelines (...) helping me, um, tick off how many serves I'm getting in each meal, something like that would be a nice bonus." (Blair, working 21 to <35 hours per week)
Organization and efficiency	• "...having the plan there it, sort of, I mean, it almost makes you more accountable (...) it really cuts out the excuses of, oh, I'm tired, we're running late, let's get a pizza." (Cora, working 21 to <35 hours per week) • "Like, it took away the decisions, decisions I had to make, I think, I had already made them, and then, I didn't need to stress about it, basically." (Blair, working 21 to <35 hours per week) • "Here's the recipe (...) adjust your shopping list (...) go through the, your supermarket of choice, get it delivered when you want, done." (Fae, working 21 to <35 hours per week)
Support and communication	• "...it would definitely be accessible across multiple devices. Um, so, you know, that, like, that everyone who's old enough and interested, in the family could contribute. (...) he wouldn't be constantly asking me every day, 'What's for tea tonight?'" (Brianna, working 35+ hours per week)
User experience — effort	
Usability	• "It's quick for meals and then the grocery, and then it comes up with a list and then you can cook it so that's what I like about it." (Fae, working 21 to <35 hours per week)
Functionality	• "I liked how it had allergies and ingredients that you liked and you disliked and how it had meal sizes, meal servings (...) you could really, like, make it for your family..." (Fae, working 21 to <35 hours per week)

Major and minor themes	Illustrative quotes
Ongoing use	• "...at least like once a week when I, because I do my grocery shopping usually once a week. So I'll probably sit down the night before and, you know, meal plan what we're going to have for that week." (Kathryn, working 1 to <21 hours per week) • "...I would do another week every couple of months on it when I was looking for inspiration." (Lana, working 35+ hours per week)

Buy-in

Early impressions of the apps appeared to be key to user buy-in and subsequent use. The alignment or fit of the app's purpose with participants' self-identified needs was important, as was the clarity of the purpose of the app. Trying to do too much or serve too many purposes was problematic.

The look and feel of the apps were also important for buy-in and used to judge credibility and trustworthiness. The esthetic quality subscale score of the uMARS aligned well with the interview data, with the more visual apps (ie, the recipe app, meal planning app, and barcode scanning app) scoring higher on the subscale. The same 3 apps ranked the highest for perceived credibility (Table 4).

The hedonic value of apps, or the pleasure associated with their use, played a role in app buy-in. Novelty was important for some participants. However, the lack of pleasure associated with the use of these apps was an issue, particularly for apps with little content. Of the 5 engagement subscale items of the uMARS, 2 relate to the hedonic value of apps, namely, the entertainment and interest qualities. Overall, the engagement subscale had the lowest scoring quality of the apps (Table 4), with the high-input and low-content apps again scoring the lowest on this domain.

Table 4. Mean (SD) uMARS scores, subjective quality score, and total score by app (n=62).

App	Value, n (%)[a]	Subscale score[b], mean (SD)				Total uMARS[c] score[b], mean (SD)	Subjective quality score[b], mean (SD)
		Engagement	Functionality	Esthetics	Credibility		
Meal planning app	35 (56.5)	3.5 (0.5)	4.2 (0.5)	4.1 (0.6)	4.0 (0.8)	3.9 (0.5)	3.0 (0.9)
Recipe manager app	32 (51.6)	3.0 (0.7)	3.7 (0.9)	3.4 (0.6)	3.8 (0.8)	3.5 (0.6)	2.7 (1.1)
Recipe app	29 (46.8)	3.7 (0.6)	4.2 (0.6)	4.3 (0.5)	4.1 (0.7)	4.1 (0.4)	3.1 (0.9)
Barcode scanning app	12 (19.4)	3.6 (0.7)	4.2 (0.6)	4.0 (0.7)	4.3 (0.8)	4.0 (0.6)	3.4 (1.0)
Family organizer app	12 (19.4)	3.4 (0.8)	3.8 (0.7)	3.8 (0.7)	3.6 (0.5)	3.6 (0.6)	2.7 (1.1)

[a]n=4 participants completed the user version of the Mobile App Rating Scale for only one app because of a lack of use of the second app.

[b]Scores range from 1 (low quality) to 5 (high quality).

[c]uMARS: user version of the Mobile App Rating Scale.

Use

More participants reported using the apps at least once in the first week than in the subsequent weeks (Multimedia Appendix 3). The barcode scanning app was used most frequently over the testing period, with at least 7 of the 12 participants allocated the app using it at least 2 to 4 times per week. Of the 12 users, 9 reported spending 1 to 5 minutes at a time on the app. There was a rapid drop-off in the use of the family organizer after the first week. There was also some decline in the use of the meal planning app over time; however, it was used for more than 5 minutes on each occasion by more than 22 of the 35 users.

Participants' descriptions of the timing and context in which they used the apps led to 2 key subthemes: in-time and when-time. In-time use of apps, while planning meals, shopping, or cooking, was purposeful or planned and undertaken to achieve a task. When-time use tended to be exploratory and took place in spare moments when it was convenient to do so. This use appeared to be for information or inspiration seeking, as opposed to functional tasks.

Another subtheme of app use was the depth to which the participants used the apps. During discussions of app features, some participants described not exploring the apps deeply enough to have knowledge of their content, features, and functionality. This may have affected participants' perception of app utility, as some even described wanting features that were already present in the apps.

Barriers

Participants described external barriers impacting their acceptance and use of the apps and their ability to incorporate the apps as new behavioral strategies for food provision. Participants reported time scarcity as a barrier to app use, whereas others suggested that they had more important priorities than using an app or changing their food provision behavior. Existing habits such as paper-based shopping lists or tried and

tested recipes were also a key barrier to participants' willingness and ability to use the apps. Habits were described as difficult to change, and the formation of new habits, such as using the apps, was seen as challenging.

User Experience

Participants' experience with the apps was organized under 2 major themes: the behavioral performance of the apps and the effort associated with use. Behavioral performance encompassed the contribution apps made to the performance of food provision behaviors, whereas effort referred to their ease of use and functionality. These aspects of the apps were weighed against one another to determine app acceptability.

Behavioral Performance

Participants found recipe content useful in providing inspiration for meals and encouraging variety. Some felt that this inspiration led to a positive dietary change. Conversely, many felt that the recipes should be better tailored to families with young children and include practical nutrition information focused on national guidelines. The barcode scanning app was found to be helpful for food selection; however, only a relatively small sample indicated a need for such support.

The organization and efficiency aspects of the meal planning app and the recipe manager app were generally found to be positive, with participants discussing planning ahead, being prepared, and feeling organized. Participants found automation features, such as automated shopping list generation, useful. Planning was reported to reduce the last-minute decision making and shopping and increase accountability. However, for those participants who already considered themselves planners, the apps were simply described as an alternative tool to use when undertaking established behaviors. A suggestion for enhancing the efficiency of the apps with automated shopping list generation was to integrate the shopping list with internet-based shopping, allowing the completion of the process from planning to purchasing.

Support and communication features were mainly found in the family organizer app, which was not well accepted by study participants due to a perceived lack of relevance to families with young children. Regardless, around half of the sample were interested in syncing between devices so that other family members could contribute to food provision–related tasks.

Effort

Ease and simplicity of use were referred to in relation to the meal planning app, barcode scanning app, and recipe app. The SUS scores aligned well with this finding, with the same 3 apps scoring above 70 (median SUS score: meal planning app 78, IQR 68-88; barcode scanning app 79, IQR 56-90; and recipe app 80, IQR 58-89), indicating a passable level of usability. The recipe manager app was also deemed to be passable (median 75, IQR 54-86) despite receiving mixed reviews during the interviews. Conversely, the usability of the family organizer app was deemed poor (median 48, IQR 34-73). Participants reported in the interviews that the meal planning app, barcode scanning app, and recipe app were particularly easy to use because they were more intuitive, self-explanatory, and required very little input from the user. Participants also spoke about the accessibility and convenience of the technology and the streamlining of processes.

The functionality subtheme of effort described the functioning of particular features of the apps. Participants liked the personalization aspects of the apps, from modifying portion sizes to filtering recipes according to dietary requirements. Some participants suggested that the inclusion of recipe importing (a feature of the recipe manager app) in the meal planning app would further enhance personalization of content. Similarly, although participants liked the automatic generation of lists, they preferred those that they could personalize or modify as per their needs. A limitation reported by participants regarding the barcode scanning app was the inability to use it while doing internet-based shopping.

Ongoing Use

Most participants reported that they would aim to use at least one of the apps periodically into the future, as required or when they had time. Those that found the apps particularly useful were clearer about their planned future use, whereas some articulated specific plans around further using the apps in different or extended ways.

Discussion

Principal Findings

This study aims to assess the feasibility of existing commercially available apps to support working parents in planning, purchasing, and preparing healthy family meals. The apps tested were found to enable in-time planning behavior, promoting organization and efficiency in food provision processes. The effort involved in using these apps had a key influence on acceptability and was weighed against the perceived benefits of the apps. The balance between these factors appeared to be key to the usefulness of these apps as tools to support food provision. The lack of family friendly recipes and nutrition content was a limitation to the utility of the apps in this sample of parents.

App Utility

Organization resulting from planning features was perceived to reduce the time burden of food provision, confirming findings from a previous review of commercial apps [14]. Planning strategies for managing food provision were found to be used by working mothers experiencing time scarcity in qualitative research [33]. These same strategies have also been associated with a higher intake of vegetables and fruits in Australian women [34]. However, planning may be challenging for those with less predictable work schedules and different family structures [35] and those who are less inclined to plan [33]. Automated and streamlined planning features, such as the generation of shopping lists from meal plans and recipes, might make these apps more widely accepted and appealing even to nonplanners.

The use of planning features in time, when food provision tasks were being undertaken, suggested that food provision apps are well placed to deliver support and content in time and context

[36]. Ecological momentary interventions (EMIs) support individual behavior in everyday life outside of research or clinical settings [13]. Evidence for EMI in the app-based nutrition space is limited [36,37], with the vast majority of research describing in-time dietary monitoring, assessment, and feedback [38-40]. The integration of these apps into daily life may be key to their ability to modify or support behavior. Furthermore, integration into daily life may overcome some of the engagement-related challenges experienced by app-based interventions in the past.

Inspiration was the main subtheme arising from discussions regarding apps with recipe content. Research investigating parental preferences for a food provision program targeting young children found that participants with higher income, older participants, and partnered participants were interested in creative cooking without recipes [41]. The sample in this study was similarly biased, perhaps explaining the use of recipe content for inspiration. The lack of family friendly recipes (ie, recipes that are acceptable to children and adults alike) may have limited parents' ability to use the recipes for anything other than inspiration.

Participants aptly suggested the need for more explicit nutrition content in the form of serve-based information linked to recipes. Previous work investigating parental preferences for an eHealth family healthy lifestyle program similarly found that parents preferred more practical nutrition information such as healthy portion sizes and recipes [42]. The barcode scanning app included in this study provided nutrition-related content and was well accepted and rated well for quality and usability by users, in line with the findings of a previous review [14]. However, according to the COM-B self-assessment, the app was not widely needed in this subset, resulting in only 12 users being allocated to the app. These findings suggest that although food provision apps may be capable of supporting the behavioral performance of food provision processes, their lack of practical nutrition content delivered in a way that suits the needs of families limits their utility in addressing diet quality in the family context.

App Acceptability

Engagement and Quality

Relatively few participants used the apps consistently and with the depth expected upon allocation. Barriers to buy-in and use included time scarcity and existing habits. Time scarcity is a common barrier to healthy food provision behaviors [11,12]. It is, therefore, unsurprising that time could also act as a barrier to the uptake of new digital solutions to food provision, especially when the use of such technology requires the formation of new habits [43]. Existing food provision habits are formed with repetition and practice and can be difficult to break, particularly in a stable environment [44]. Previous research has similarly found that time and habits are key barriers to food provision behaviors and the uptake of a meal planning app targeting low-income parents [45]. Aligning app use with everyday food provision tasks (such as meal planning), automation of key features, and integration with services such as internet-based shopping may alleviate the time burden of app use itself. Positioning food provision apps as tools to support the maintenance of healthy habits during times of stress or disruption may reduce the need for new habit formation.

Consistent with previous work [14,46], engagement quality was the lowest scoring uMARS subscale, particularly in those apps with minimal existing content. This is concerning, as user engagement is a major challenge to the efficacy and longevity of mHealth interventions [9]. The novelty of the apps was positive, as it is thought to play an early role in the hedonic value of technology [43]. However, familiarity with technology tends to reduce the pleasure derived from its novelty [43]. This suggests a need for other qualities to promote ongoing engagement. As the enjoyment or pleasure associated with the use of technology has been shown to be important to the usability, acceptance, and use of apps [43,47], enhancing this aspect should be a key consideration in the future. Features such as gamification can achieve this [48]. Providing an adequate onboarding experience highlighting the app's purpose and functionality may also be helpful in promoting a better depth of engagement.

Usability and Functionality

Participants in this study weighed the behavioral performance of the apps against the effort required to make use of them, with effort reflecting aspects of usability and functionality. According to the consumer version of the Unified Theory of Acceptance and Use of Technology, the effort involved in technology use is thought to be a key predictor of intention to use, acceptance, and actual use of technology [43]. Furthermore, the level of effort or the process involved in using technology may be more important to women [43], who made up the majority of participants in this study. The greater acceptance of the apps requiring less user input (such as the recipe app and the barcode scanning app) suggests the need for careful consideration of the balance between effort and behavioral performance in future app development. This finding strengthens the case for automated features and integration of apps with daily life, which has been shown to be a high priority for both digital health experts and consumers in addressing the usability of health-related apps [47]. The generation of shopping lists from meal plans, entry of grocery items using a barcode scanner, and integration with internet-based shopping might improve app usability.

Context and Need

The parents in this study identified a need for information in the form of healthy recipes and meal ideas and for ways to reduce the time burden of food provision rather than for support with food choice or food preparation skills. This may reflect the higher income and education level of the sample, which has been previously associated with greater food and nutrition knowledge, skills, and confidence [49,50]. A discrete choice experiment showed that older parents, parents with higher income, and partnered parents had a preference for meal planning and time-saving strategies, whereas younger parents, parents with lower income, and single parents preferred support with healthy cooking and nutrition [41]. Future research in households where needs are different could consider apps and app features that were less represented in this work (eg, apps focused on nutrition information and food preparation skills).

Strengths and Limitations

Although this study does not assess dietary behavior change resulting from the use of these mobile apps, it provides early evidence to support future app development and testing. The strength of this study was its mixed methods approach, including the allocation of apps based on need and the incorporation of rich qualitative data triangulated with quantitative findings. However, despite allocating apps according to need, they were not always deemed relevant or suitable. Parents testing the family organizer app felt that it was not relevant to families of young children; therefore, the feedback regarding this app should be interpreted with caution. There were also limitations with regard to the sample population, which may limit the generalizability of the findings. The sample was typically of high socioeconomic status, which may have led to homogeneity in the results. However, efforts were made to incorporate the voices of single parents and parents of lower socioeconomic status. Irrespective of this sample bias, the vegetable, fruit, and discretionary choice intake of the study sample reflected the eating habits of the broader Australian population [51,52]. Therefore, the sample in this study would benefit from food provision–related support as much as the broader Australian population.

Implications for Practice and Future Research

The behavior change potential of food provision apps may lie in their ability to be integrated into everyday life, promoting healthy food provision in time and context. Meal planning apps with automated planning and shopping list preparation and integration with internet-based shopping and between users may provide a nexus between dietary guidelines and healthy food provisioning and enable planning behavior in those less inclined to plan. Before more rigorous efficacy testing, future research should endeavor to strengthen the behavior change content of these apps by including features directly addressing time scarcity and parents' need for healthy recipes and meal ideas. However, other digital tools that may be used to achieve similar goals (such as recipe websites, grocery shopping websites, and social media) are also worth considering.

Conclusions

This study has provided insights into the role of mobile apps in supporting parents to achieve healthy food provision in a family context. Meal planning apps and features promoting organization present feasible, time-saving solutions to support healthy food provision practices. However, the time burden of app use may outweigh the time saved in the food provision process. A balance must be achieved between effort and outcome to improve the usability and usefulness of these apps. To progress in this area, attention must be paid to enhancing app automation and integration as well as recipe and nutrition content to support healthy food provision behavior in time.

Acknowledgments

This research was funded by the Flinders Foundation. CEM was supported by an Australian Government Research Training Program Scholarship; the National Health and Medical Research Council postgraduate scholarship (1114194), a top-up scholarship funded by Flinders University; the National Health and Medical Research Council Centre for Research Excellence in The Early Prevention of Obesity in Childhood (1101675); and a King and Amy O'Malley Trust Postgraduate Scholarship. RAL and RKG are researchers within the National Health and Medical Research Council Centre for Research Excellence in The Early Prevention of Obesity in Childhood (1101675).

Authors' Contributions

CEM led the study design under the supervision of RAL, TPW, and RKG and in consultation with IP and AJM. CEM conducted the research, with all coauthors contributing to the interpretation of the results. CEM drafted the manuscript, and all coauthors reviewed drafts and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of apps tested.
[DOCX File , 24 KB - mhealth_v9i5e22990_app1.docx]

Multimedia Appendix 2
Mapping of apps to COM-B (Capability, Opportunity, Motivation, and Behavior) items.
[DOCX File , 24 KB - mhealth_v9i5e22990_app2.docx]

Multimedia Appendix 3
Frequency and duration of app use.
[DOCX File , 26 KB - mhealth_v9i5e22990_app3.docx]

References

1. Australian Institute of Health and Welfare. A picture of overweight and obesity in Australia 2017. Canberra, Australia: Australian Government; 2017. URL: https://www.aihw.gov.au/reports/overweight-obesity/a-picture-of-overweight-and-obesity-in-australia/summary [accessed 2020-07-13]
2. World Health Organization. Noncommunicable diseases country profiles 2018. Geneva, Switzerland: World Health Organization; 2018. URL: https://www.who.int/nmh/publications/ncd-profiles-2018/en/ [accessed 2020-05-19]
3. Australian Institute of Health and Welfare. Nutrition across the life stages. Canberra, Australia: Australian Government; 2018. URL: https://www.aihw.gov.au/reports/food-nutrition/nutrition-across-the-life-stages/summary [accessed 2020-07-06]
4. Moubarac J, Batal M, Louzada M, Martinez Steele E, Monteiro C. Consumption of ultra-processed foods predicts diet quality in Canada. Appetite 2017 Jan;108:512-520. [doi: 10.1016/j.appet.2016.11.006]
5. US Department of Health and Human Services, US Department of Agriculture. Dietary Guidelines for Americans 2015-2020, 8th Edition. United States; 2015. URL: https://health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf [accessed 2020-06-26]
6. Spence AC, Campbell KJ, Lioret S, McNaughton SA. Early Childhood Vegetable, Fruit, and Discretionary Food Intakes Do Not Meet Dietary Guidelines, but Do Show Socioeconomic Differences and Tracking over Time. Journal of the Academy of Nutrition and Dietetics 2018 Sep;118(9):1634-1643.e1. [doi: 10.1016/j.jand.2017.12.009]
7. Australian Bureau of Statistics. 0.55. 6224.0.55.001 - Labour Force, Australia: Labour Force Status and Other Characteristics of Families. Canberra, Australia: Australian Government; 2019. URL: https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-status-families/latest-release [accessed 2020-07-06]
8. Organisation for Economic Cooperation and Development. Connecting People with Jobs: Key Issues for Raising Labour Market Participation in Australia. Paris, France: OECD publishing; 2017. URL: https://www.oecd-ilibrary.org/employment/connecting-people-with-jobs-key-issues-for-raising-labour-market-participation-in-australia_9789264269637-en;jsessionid=LeiGipTcOvYEaAopSaLaGT5u.ip-10-240-5-59 [accessed 2020-07-06]
9. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Sci 2011 Apr 23;6(42):1-11. [doi: 10.1186/1748-5908-6-42]
10. Johnson BJ, Zarnowiecki D, Hendrie GA, Mauch CE, Golley RK. How to reduce parental provision of unhealthy foods to 3- to 8-year-old children in the home environment? A systematic review utilizing the Behaviour Change Wheel framework. Obesity Reviews 2018 Aug 09;19(10):1359-1370. [doi: 10.1111/obr.12702]
11. Jabs J, Devine CM. Time scarcity and food choices: An overview. Appetite 2006 Sep;47(2):196-204. [doi: 10.1016/j.appet.2006.02.014]
12. Venn D, Strazdins L. Your money or your time? How both types of scarcity matter to physical activity and healthy eating. Social Science & Medicine 2017 Jan;172:98-106. [doi: 10.1016/j.socscimed.2016.10.023]
13. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
14. Mauch CE, Wycherley TP, Laws RA, Johnson BJ, Bell LK, Golley RK. Mobile Apps to Support Healthy Family Food Provision: Systematic Assessment of Popular, Commercially Available Apps. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11867. [doi: 10.2196/11867]
15. Crawford DA, Ball K, Cleland VJ, Campbell KJ, Timperio AF, Abbott G, et al. Home and neighbourhood correlates of BMI among children living in socioeconomically disadvantaged neighbourhoods. Br J Nutr 2011 Aug 09;107(7):1028-1036. [doi: 10.1017/s0007114511003801]
16. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126]
17. 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. [doi: 10.2196/jmir.4055]
18. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework. PLoS ONE 2016 Mar 15;11(3):e0150205. [doi: 10.1371/journal.pone.0150205]
19. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: A Mobile Phone App to Enable Consumers to Make Healthier Food Choices and Crowdsourcing of National Food Composition Data. JMIR mHealth uHealth 2014 Aug 21;2(3):e37. [doi: 10.2196/mhealth.3230]
20. Bell LK, Golley RK, Magarey AM. A short food-group-based dietary questionnaire is reliable and valid for assessing toddlers' dietary risk in relatively advantaged samples. Br J Nutr 2014 Jun 02;112(4):627-637. [doi: 10.1017/s0007114514001184]
21. Litterbach E, Russell CG, Taki S, Denney-Wilson E, Campbell KJ, Laws RA. Factors Influencing Engagement and Behavioral Determinants of Infant Feeding in an mHealth Program: Qualitative Evaluation of the Growing Healthy Program. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e196. [doi: 10.2196/mhealth.8515]

22. Ball K, Mouchacca J, Jackson M. The feasibility and appeal of mobile 'apps' for supporting healthy food purchasing and consumption among socioeconomically disadvantaged women: a pilot study. Health Promot J Austr 2014 Sep 04;25(2):79-82. [doi: 10.1071/he13096]
23. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of Smartphone Applications for Nutrition and Physical Activity Behavior Change. JMIR Res Protoc 2012 Aug 22;1(2):e9. [doi: 10.2196/resprot.2205]
24. Mummah SA, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 8;13(1):90-102. [doi: 10.1186/s12966-016-0400-z]
25. Michie S, Atkins L, West R. The Behaviour Change Wheel: A guide to designing interventions. Great Britain: Silverback Publishing; 2014.
26. Hendrie GA, Rebuli MA, Golley RK. Reliability and relative validity of a diet index score for adults derived from a self-reported short food survey. Nutr Diet 2016 Sep 14;74(3):291-297. [doi: 10.1111/1747-0080.12303]
27. Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry 1996;189(194):4-7. [doi: 10.1201/9781498710411-35]
28. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
29. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care 2007 Sep 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042]
31. Hendrie GA, Rebuli MA, Golley RK, Noakes M. Adjustment Factors Can Improve Estimates of Food Group Intake Assessed Using a Short Dietary Assessment Instrument. Journal of the Academy of Nutrition and Dietetics 2018 Oct;118(10):1864-1873. [doi: 10.1016/j.jand.2018.02.018]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Jabs J, Devine CM, Bisogni CA, Farrell TJ, Jastran M, Wethington E. Trying to Find the Quickest Way: Employed Mothers' Constructions of Time for Food. Journal of Nutrition Education and Behavior 2007 Jan;39(1):18-25. [doi: 10.1016/j.jneb.2006.08.011]
34. Crawford D, Ball K, Mishra G, Salmon J, Timperio A. Which food-related behaviours are associated with healthier intakes of fruits and vegetables among women? Public Health Nutr 2007 Mar 01;10(3):256-265. [doi: 10.1017/s1368980007246798]
35. Devine CM, Farrell TJ, Blake CE, Jastran M, Wethington E, Bisogni CA. Work Conditions and the Food Choice Coping Strategies of Employed Parents. Journal of Nutrition Education and Behavior 2009 Sep;41(5):365-370. [doi: 10.1016/j.jneb.2009.01.007]
36. Boh B, Lemmens LHJM, Jansen A, Nederkoorn C, Kerkhofs V, Spanakis G, et al. An Ecological Momentary Intervention for weight loss and healthy eating via smartphone and Internet: study protocol for a randomised controlled trial. Trials 2016 Mar 22;17(1):1-12. [doi: 10.1186/s13063-016-1280-x]
37. Brookie KL, Mainvil LA, Carr AC, Vissers MC, Conner TS. The development and effectiveness of an ecological momentary intervention to increase daily fruit and vegetable consumption in low-consuming young adults. Appetite 2017 Jan;108:32-41. [doi: 10.1016/j.appet.2016.09.015]
38. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A Mobile Ecological Momentary Assessment Tool (devilSPARC) for Nutrition and Physical Activity Behaviors in College Students: A Validation Study. J Med Internet Res 2016 Jul 27;18(7):e209. [doi: 10.2196/jmir.5969]
39. Poelman MP, van Lenthe FJ, Scheider S, Kamphuis CB. A Smartphone App Combining Global Positioning System Data and Ecological Momentary Assessment to Track Individual Food Environment Exposure, Food Purchases, and Food Consumption: Protocol for the Observational FoodTrack Study. JMIR Res Protoc 2020 Jan 28;9(1):e15283. [doi: 10.2196/15283]
40. Comulada WS, Swendeman D, Koussa MK, Mindry D, Medich M, Estrin D, et al. Adherence to self-monitoring healthy lifestyle behaviours through mobile phone-based ecological momentary assessments and photographic food records over 6 months in mostly ethnic minority mothers. Public Health Nutr 2017 Dec 04;21(4):679-688. [doi: 10.1017/s1368980017003044]
41. Virudachalam S, Chung PJ, Faerber JA, Pian TM, Thomas K, Feudtner C. Quantifying parental preferences for interventions designed to improve home food preparation and home food environments during early childhood. Appetite 2016 Mar;98:115-124. [doi: 10.1016/j.appet.2015.11.007]
42. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition Interventions for Prevention and Management of Childhood Obesity: What Do Parents Want from an eHealth Program? Nutrients 2015 Dec 15;7(12):10469-10479. [doi: 10.3390/nu7125546]
43. Venkatesh V, Thong JY, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]

44. Ouellette JA, Wood W. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. Psychological Bulletin 1998 Jul;124(1):54-74. [doi: 10.1037/0033-2909.124.1.54]
45. Garvin TM, Chiappone A, Boyd L, Stern K, Panichelli J, Edwards Hall LA, et al. Cooking Matters Mobile Application: a meal planning and preparation tool for low-income parents. Public Health Nutr 2019 May 14;22(12):2220-2227. [doi: 10.1017/s1368980019001101]
46. Flaherty S, McCarthy M, Collins A, McAuliffe F. Can existing mobile apps support healthier food purchasing behaviour? Content analysis of nutrition content, behaviour change theory and user quality integration. Public Health Nutr 2017 Oct 30;21(2):288-298. [doi: 10.1017/s1368980017002889]
47. Liew MS, Zhang J, See J, Ong YL. Usability Challenges for Health and Wellness Mobile Apps: Mixed-Methods Study Among mHealth Experts and Consumers. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e12160. [doi: 10.2196/12160]
48. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS ONE 2017 Mar 31;12(3):e0173403. [doi: 10.1371/journal.pone.0173403]
49. Burton M, Reid M, Worsley A, Mavondo F. Food skills confidence and household gatekeepers' dietary practices. Appetite 2017 Jan;108:183-190. [doi: 10.1016/j.appet.2016.09.033]
50. McLeod ER, Campbell KJ, Hesketh KD. Nutrition Knowledge: A Mediator between Socioeconomic Position and Diet Quality in Australian First-Time Mothers. Journal of the American Dietetic Association 2011 May;111(5):696-704. [doi: 10.1016/j.jada.2011.02.011]
51. Australian Bureau of Statistics. 4364.0.55.007 - Australian Health Survey: Nutrition First Results - Foods and Nutrients, 2011-12. 2011. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/australian-health-survey-nutrition-first-results-foods-and-nutrients/latest-release [accessed 2020-07-06]
52. Fayet-Moore F, McConnell A, Cassettari T, Tuck K, Petocz P, Kim J. Discretionary intake among Australian adults: prevalence of intake, top food groups, time of consumption and its association with sociodemographic, lifestyle and adiposity measures. Public Health Nutr 2019 Jan 25;22(09):1576-1589. [doi: 10.1017/s1368980018003361]

Abbreviations

COM-B: Capability, Opportunity, Motivation, and Behavior
EMI: ecological momentary intervention
NCD: noncommunicable disease
SUS: System Usability Scale
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 28.07.20; peer-reviewed by E Hu, M allman-Farinelli; comments to author 18.09.20; revised version received 26.10.20; accepted 29.03.21; published 07.05.21.

Please cite as:
Mauch CE, Laws RA, Prichard I, Maeder AJ, Wycherley TP, Golley RK
Commercially Available Apps to Support Healthy Family Meals: User Testing of App Utility, Acceptability, and Engagement
JMIR Mhealth Uhealth 2021;9(5):e22990
URL: https://mhealth.jmir.org/2021/5/e22990
doi:10.2196/22990
PMID:33960951

Original Paper

Acceptability of the Pregnancy, Exercise, and Nutrition Research Study With Smartphone App Support (PEARS) and the Use of Mobile Health in a Mixed Lifestyle Intervention by Pregnant Obese and Overweight Women: Secondary Analysis of a Randomized Controlled Trial

Ellen M Greene[1], BSc; Eileen C O'Brien[1], BSc, PhD; Maria A Kennelly[1], PhD, MB, BCh, BAO, MRCPI; Orna A O'Brien[1], BSc, MSc; Karen L Lindsay[1,2], RD, PhD; Fionnuala M McAuliffe[1], MD, FRCOG, FRCPI, Hon FACOG

[1]UCD Perinatal Research Centre, School of Medicine, University College Dublin, National Maternity Hospital, Dublin, Ireland

[2]Department of Pediatrics, University of California, Irvine, CA, United States

Corresponding Author:
Fionnuala M McAuliffe, MD, FRCOG, FRCPI, Hon FACOG
UCD Perinatal Research Centre
School of Medicine
University College Dublin, National Maternity Hospital
UCD Perinatal Research Centre, School of Medicine
University College Dublin, National Maternity Hospital
Dublin, Dublin 2
Ireland
Phone: 353 637 3216
Email: fionnuala.mcauliffe@ucd.ie

Abstract

Background: Dietary interventions can improve pregnancy outcomes among women with increased BMI. Although the interest in mobile health interventions is growing, little is known about the acceptability of smartphone apps to support lifestyle interventions in such a cohort.

Objective: We aimed to assess the acceptability of the pregnancy, exercise, and nutrition research study with smartphone app support (PEARS) and the use of mobile health in a mixed lifestyle intervention delivered to overweight and obese pregnant women.

Methods: PEARS was a randomized controlled trial of a low glycemic index dietary intervention with exercise prescription and a smartphone app, which was delivered to pregnant women who were overweight or obese. Acceptability questionnaires were completed by the intervention group at 28 weeks of gestation (n=149) and at postintervention (n=123). Maternal characteristics were recorded (ie, age, ethnicity, BMI, socioeconomic status). Associations between maternal characteristics and acceptability of the intervention and app were analyzed using two-tailed *t* tests, Mann-Whitney *U* tests, chi-square test, and logistic regression. One-on-one semistructured interviews were conducted with a subcohort of the intervention participants (n=28) at 34 weeks of gestation, in which the participants shared their experiences of the PEARS intervention.

Results: The intervention was generally accepted, with respondents agreeing that the diet was easy to follow (98/148, 68.5%), enjoyable (106/148, 74.1%), and affordable (110/148, 76.9%). Qualitative and quantitative results were consistent with each another, both demonstrating that app acceptability was high. The participants agreed that the app was enjoyable (96/120, 80.0%) and easy to use (116/119, 97.5%). Compared to those with tertiary education, those with lower education levels were more likely to enjoy the dietary changes (*P*=.04). Enjoyment of the app was associated with disadvantaged neighborhood deprivation index (*P*=.01) and higher BMI (*P*=.03).

Conclusions: The PEARS intervention and use of a supportive smartphone app were accepted by pregnant women, particularly by those from vulnerable subgroups of this population.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 29316280; https://www.isrctn.com/ISRCTN29316280

(JMIR Mhealth Uhealth 2021;9(5):e17189) doi:10.2196/17189

KEYWORDS

pregnancy; mHealth; nutrition; lifestyle; acceptability; app; mobile phone

Introduction

Mobile Health as a Support in Lifestyle Interventions

Mobile health (mHealth) is the use of mobile technology such as smartphones, personal digital assistants, or other wireless devices as a tool for medical or public health care purposes [1]. mHealth technology is now growing in popularity as a component of research interventions for both pregnant and nonpregnant cohorts. mHealth as part of mixed lifestyle interventions has been successful in promoting weight loss and increasing physical activity levels in nonpregnant normal weight, overweight, and obese individuals [2,3]. The use of mobile devices in such interventions has been deemed acceptable and found to be highly compliant, particularly among overweight individuals [3]. Furthermore, mHealth-supported diet and exercise interventions in pregnancy have been deemed effective and acceptable among overweight and obese cohorts [4-8]. However, the majority of studies used mHealth in the form of text messaging, websites, and fitness trackers [4-8]. There is a paucity of data on the acceptability of the use of mHealth in the form of a smartphone app specifically to support a mixed lifestyle intervention in pregnancy.

Use of mHealth Among Hard-to-Reach Populations

Through the use of mHealth, there is increased potential to provide support to those subgroups of the population who experience barriers in obtaining health care information, such as low education level and low income or residing in difficult-to-reach geographical locations. Lower socioeconomic groups have relatively higher levels of ill health, fewer resources, and are more likely to report having chronic health conditions such as diabetes [9,10]. If designed accordingly, mobile technology could be used to present supplementary dietary and physical activity information using easy-to-understand language, which is supported by graphics. A smartphone app is a convenient and economical tool for communication as well as a potential additional support for those with few resources to obtain health care information. However, a review of the use of pregnancy apps highlighted that women with low incomes were among the groups with the lowest rate of pregnancy app uptake [11]. In order to ensure that smartphone apps communicating diet and lifestyle information are reached to lower socioeconomic groups, lower-income populations should be considered in the app development and the acceptability of reliable pregnancy apps should be examined in such groups.

Pregnancy, Exercise, and Nutrition Research Study With Smartphone App Support

The pregnancy, exercise, and nutrition research study with smartphone app support (PEARS) used mHealth in the form of a smartphone app to support a diet and exercise intervention aimed to reduce the incidence of gestational diabetes mellitus (GDM) [12]. GDM is a metabolic condition of glucose intolerance in pregnancy with negative implications for both the mother and the fetus, for which a high prepregnancy BMI is a risk factor [13-15]. Currently, 53% of the female Irish population are overweight or obese [10], highlighting the need for interventions targeting at-risk groups to reduce the incidence of GDM. Dietary and lifestyle approaches are effective in improving outcomes [16-18], and several diet and exercise interventions carried out in pregnant cohorts, thus far, have been deemed acceptable by participants [19,20]. However, only few interventions aiming to reduce the incidence of GDM in at-risk groups have incorporated a smartphone app into the intervention design and therefore, the acceptability of a smartphone app used in this context is unknown. As pregnancy app uptake differs across various population groups [11], maternal demographics should be considered when examining smartphone app acceptability.

Aims and Objectives of This Study

We aimed to quantitatively and qualitatively examine participant acceptability and demographics associated with the acceptability of the (1) PEARS intervention and (2) PEARS diet and exercise smartphone app.

Methods

Study Design

This study was a secondary data analysis of pregnant women originally recruited as part of the PEARS between 2013 and 2016 at the National Maternity Hospital, Dublin, Ireland. The detailed methodology and results of the PEARS study have been previously published [12,21]. In brief, PEARS was a randomized controlled trial that evaluated the effect of a "healthy lifestyle package" on the incidence of GDM in 565 pregnant women who were overweight or obese. The intervention group (n=278) received low glycemic index (GI) dietary advice, a daily exercise prescription, and a study-specific smartphone app as well as standard obstetric care throughout pregnancy. The control group received standard obstetric care only. The primary outcome was the incidence of GDM among participants at 28 weeks of gestation for which no difference was observed between the groups [12]. Differences were noted between intervention and control groups in gestational weight gain and in the delivery rates of large-for-gestational-age infants [12].

Ethical Approval

Institutional ethical approval for the PEARS study was granted by The National Maternity Hospital Ethics committee in October 2012.

Patient Selection

Pregnant women were recruited from the National Maternity Hospital. The predefined, published protocol was followed [21]. Eligibility criteria included possession of a smartphone, age of 18-45 years, between 10 and 16 weeks of gestation, early

pregnancy BMI≥25 kg/m² and ≤39.9 kg/m², singleton pregnancy, and an adequate level of understanding of the English language to give informed consent.

Assessment of Maternal Characteristics

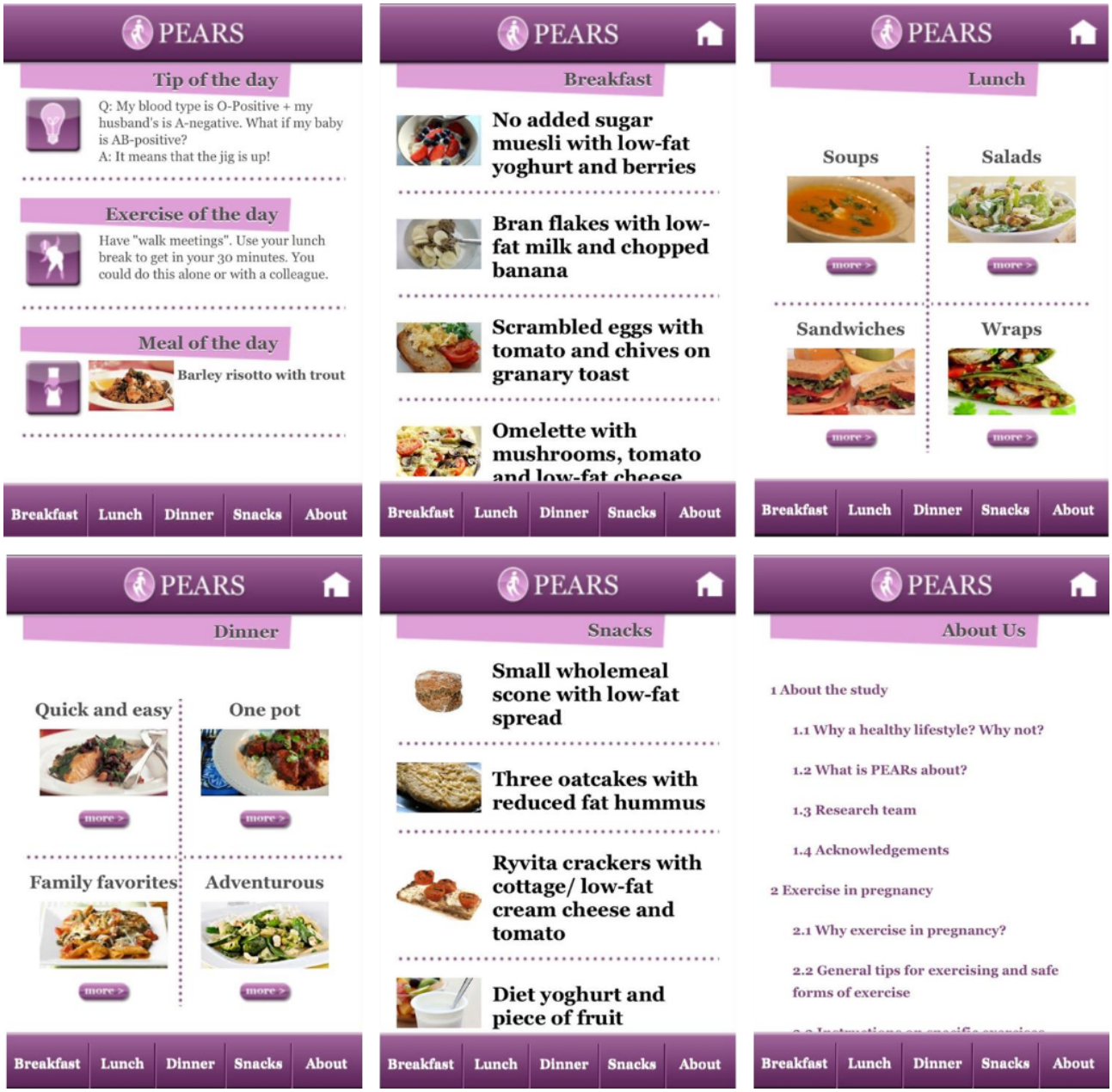
Maternal weight and height were measured at enrolment and BMI was calculated (kg/m²) and the women were categorized as overweight (25.0-29.9 kg/m²) or obese (≥30 kg/m²). Age, parity, and ethnicity were collected from medical charts. Education level was self-reported by participants and they were classified as having achieved tertiary education or not. Neighborhood deprivation was assessed according to the Pobal Haase-Pratschke (Pobal HP) Deprivation Index [22] using participants' home addresses and was categorized as advantaged (score>0) or disadvantaged (score<0).

The PEARS Smartphone App

App Development

The smartphone app provided to the intervention group was a study-specific app designed to support the diet and physical activity information and advice provided by clinical research personnel during an in-person education session after enrolment (Figure 1). The smartphone app was designed by a multidisciplinary team, including a dietician, obstetrician, food behavior specialist, and an app design company. Focus groups were carried out among pregnant women during the development of the app to inform the app's content. The focus groups were conducted by a research obstetrician and included 10 randomly selected pregnant women with an overweight or obese BMI. Information on these women's knowledge of diet and exercise as well as their needs and wants in a lifestyle pregnancy app were used to develop the app's design and content. The app's text was developed for a reading age of 12 years to allow those with lower literacy levels to engage with the app. It was also developed in an easy-to-use format to allow for digital literacy.

Figure 1. Screenshots of the pregnancy, exercise, and nutrition research study with smartphone app support (PEARS) app.



App Usage Recommendations

The app was downloaded by PEARS participants at approximately 16 weeks of gestation and was password-protected so that only participants who were in the intervention group could access it. Participants were encouraged by researchers to use the app every day and they were encouraged to set a recurring alarm on their phones as a reminder. The app had 3 main sections: a *home* page, an *about* page, and a *database of low GI meal ideas and recipes* developed by a research dietician. The home page contained a "Tip of the Day," which gave a piece of general advice for pregnancy and an "Exercise of the Day" and "Meal of the Day," which supported the low-GI diet and exercise advice given at the education session. The *about* page contained additional information about the PEARS study and physical activity in pregnancy. Participant app use was recorded by the app's software from the baseline visit to the woman's delivery date. App usage data were downloaded and total instances of app use per person were calculated. An "instance" of use of the app was defined as a 15-minute period in which the app was opened once or more.

Participant Feedback Questionnaires

Participants were included if they returned a completed feedback questionnaire on intervention acceptability (Multimedia Appendix 1) at 34 weeks of gestation, which assessed compliance with and acceptability of each aspect of the intervention using a Likert-type scale or questions where an option was given to "choose all that apply." The dietary component of the questionnaire was based on a questionnaire used in the ROLO study [16], which was originally based on questions used in a low GI intervention study by Moses et al [23]. The physical activity questions were based on the Motives for Physical Activity Measure questionnaire [24]. Upon completion of the intervention, a smartphone app acceptability

questionnaire, which assessed the enjoyment and acceptability of the app, was completed by the participants (Multimedia Appendix 2). This questionnaire was self-designed, as at the time of study development, there were no existing validated questionnaires examining the use of smartphone apps to deliver a lifestyle intervention during pregnancy.

Statistical Analyses

Statistical analyses of quantitative data were carried out using SPSS Statistics v.20 (IBM Corp). Descriptions of maternal characteristics and questionnaire responses were presented as mean (SD) or n (%) for continuous and categorical data, respectively. The normality of the continuous variables was assessed visually using histograms. Bivariate analysis was initially carried out to test for associations between maternal characteristics and responses to individual questions in the 2 questionnaires (independent samples two-tailed *t* tests, Mann-Whitney *U* tests, and chi-square analyses). Results were presented as mean (SD), unless otherwise stated. Variables that were significantly associated in bivariate analysis were further analyzed using logistic regression, controlling for age, BMI, socioeconomic status, education level, and parity. The analysis for each question was conducted separately. Mann-Whitney *U* tests were carried out to test for associations between total instances of app use (objectively obtained from the app) and app acceptability (based on questionnaire responses). Where significant associations were obtained, logistic regression analysis was carried out, controlling for smoking, age, socioeconomic status, education level, BMI, parity, and ethnicity.

Qualitative Interviews

A subset of the PEARS cohort was invited to partake in semistructured one-on-one interviews, in which participants were asked to describe their experiences of the PEARS study and app. A subset of 30 women was to be purposely sampled to recruit a balance of nulliparous (15/30, 50%) and multiparous (15/30, 50%) women who had either obtained a third-level qualification (15/30, 50%) or whose highest education achievement was second-level education or below (15/30, 50%). The interview occurred at 34 weeks of gestation. The interviews were voice-recorded and transcribed by a commercial transcription service. The transcribed interviews were anonymized and thematically analyzed. Participants were assigned pseudonyms for reporting purposes.

Results

Maternal Characteristics and Responses to Questionnaires

Of the 278 participants in the intervention group, 149 participants completed the PEARS study feedback form (53.6% response rate) and 123 completed the Smartphone App Evaluation Questionnaire (44.2% response rate). There were no differences in the maternal demographics between those who responded to the questionnaires and those who did not (Multimedia Appendix 3). The characteristics of the respondents are outlined below (Table 1). Responses to both questionnaires are presented in Table 2.

Table 1. Characteristics of the participants who completed the questionnaires.

Characteristic	PEARS[a] study feedback form		Smartphone app evaluation questionnaire	
	n	Values	n	Values
Age (years), mean (SD)	147	32.84 (4.61)	123	32.78 (4.45)
Early-pregnancy weight (kg), mean (SD)	149	78.61 (10.67)	123	78.99 (10.99)
Early-pregnancy BMI (kg/m²), mean (SD)	149	29.21 (3.32)	123	29.11(3.28)
Overweight (BMI 25-29.9 kg/m²), n (%)	149	104 (69.8)	123	90 (73.2)
Obese (BMI ≥30 kg/m²), n (%)	149	45 (30.2)	123	33 (26.8)
Neighborhood deprivation index (Pobal Haase-Pratschke index), mean (SD)	149	5.46 (11.22)	123	5.4 (11.31)
Advantaged, n (%)[b]	149	105 (70.5)	123	84 (68.3)
Disadvantaged, n (%)[c]	149	44 (29.5)	123	39(31.7)
Achieved third-level education, n (%)	145	88 (60.7)	107	63 (58.9)
Smoking in early pregnancy, n (%)	137	8 (5.8)	108	3 (2.8)
Parity, median (IQR)	147	0 (0-1)	122	1 (0-1)
Multiparous, n (%)	147	73 (49.7)	122	62 (50.8)
White-Irish ethnicity, n (%)	148	123 (83.1)	121	100 (82.6)

[a]PEARS: pregnancy, exercise, and nutrition research study with smartphone app support.

[b]Advantaged is indicative of a Pobal Haase-Pratschke index >0.

[c]Disadvantaged is indicative of a Pobal Haase-Pratschke index ≤0.

Table 2. Data of the responses on the acceptability of the intervention and smartphone app.

Question	Answer	Values, n (%)
Intervention acceptability		
I have followed the recommended diet (n=153)	Compliant (always or mostly)	101 (67.8)
It was easy to follow the diet recommended during the study (n=148)	Agree[a]	98 (68.5)
I enjoyed the dietary changes I made (n=148)	Agree	106 (74.1)
The changes I made did not increase my weekly grocery bill (n=148)	Agree	110 (76.9)
My family was happy with the changes I made to my diet (n=148)	Agree	98 (68.5)
I felt I had enough energy while on the diet (n=149)	Agree	113 (78.5)
I enjoyed eating a wide variety of foods in my eating plan (n=148)	Agree	118 (82.5)
I followed the exercise prescription (n=153)	Compliant (regularly)	67 (45.3)
I adhered to the exercise that was prescribed to me because (Choose all that apply) (n=141)		
	I was told to by the research team	28 (19.9)
	I knew it was beneficial for me and the pregnancy	123 (87.2)
	I was influenced by the app and other social media	8 (5.7)
	I wanted to feel better about myself	67 (47.5)
	It was unavoidable	15 (10.6)
	I had good support from family and friends	29 (20.6)
I was unable to perform exercise because of the following (Choose all that apply) (n=83)		
	Lack of time	46 (55)
	Lack of facilities	1 (1)
	Lack of support	1 (1)
	Weather	31 (37)
	Prohibition by family, friends, doctors	11 (13)
	Didn't care to do it	10 (12)
	Lack of guidance by research team	0 (0)
	Lack of understanding of the exercise	0 (0)
	Worried in case it wasn't safe in pregnancy	7 (9)
	I felt it was ineffective and pointless	0 (0)
App acceptability		
The PEARS[b] app was enjoyable to use (n=120)		96 (80.0)
The PEARS app was easy to use (n=119)		116 (97.5)
The PEARS app was written in language that was easy to understand (n=120)		120 (100.0)
The PEARS app was attractively presented (n=123)		113 (91.9)
The PEARS app was graphically helpful (n=118)		99 (83.9)
The PEARS app was useful (n=122)		112 (91.8)
The PEARS app made me think about my diet (n=120)		107 (89.2)
I found the "Tip of the Day" function		
	Useful (n=120)	103 (85.8)
	Practical (n=120)	96 (80.0)
	Motivating (n=118)	87 (73.7)

Question	Answer	Values, n (%)
	Motivated me to eat well (n=118)	87 (73.7)
	Motivated me to be active (n=118)	79 (66.9)
	Was helpful in planning meals (n=118)	71 (60.2)
I found the "Exercise of the Day" function		
	Useful (n=121)	91 (75.2)
	Practical (n=119)	85 (71.4)
	Motivating (n=120)	83 (69.2)
I found the "Meal of the Day" function		
	Useful (n=122)	99 (81.1)
	Practical (n=118)	90 (76.3)
	Motivating (n=120)	88 (73.3)
	Agree that the meals on the app were no more expensive than meals I made before I began the study (n=121)	89 (73.6)
There was () detail provided on the app (n=122)		
	Too little	28 (23.0)
	Enough	97 (79.5)
	Too much	0 (0.0)
Which of the following enticed you to use the app regularly (Choose all that apply) (n=123)		
	Looks good	14 (11.4)
	Easy to use	76 (61.8)
	Readily available	67 (54.5)
	Exactly what I need	12 (9.8)
	Answers my questions on diet and exercise	26 (21.1)
	Meal of the Day function	49 (39.8)
	Exercise of the Day function	23 (18.7)
	Tip of the Day function	49 (39.8)
How many times per week did you use the app? (n=123)	Regularly	79 (64.2)
Would you recommend this app to a friend? (n=123)	Yes	116 (94.3)
Would you use this app if you were pregnant again? (n=123)	Yes	115 (93.5)

[a]Agree: the sum of the responses "agree" and "strongly agree."

[b]PEARS: pregnancy, exercise, and nutrition research study with smartphone app support.

App Usage

The median weeks of app use among participants was 20 (IQR 11-24) weeks. Over this time period, the median of total instances of app use for participants was 24 (IQR 8-72). The median for instances of app use per week among participants was 1.75 (IQR 0.71-3.49).

Bivariate Analysis

Intervention Acceptability

The mean neighborhood deprivation scores in the group who found the recommended diet easy to follow were significantly lower than those in the group who did not (4.46 [SD 11.25] vs 8.58 [SD 10.95], respectively; *P*=.04). Those who did not achieve tertiary education were more likely to agree that they enjoyed the dietary changes than those who achieved tertiary education (48/54, 89% (SD 32%) vs 55/86, 64% (SD 48%), respectively; *P*=.002). Those who felt that dietary changes did not increase their grocery bill were older than those who did (33.42 [SD 4.68] years vs 31.45 [SD 3.92] years, respectively; *P*=.03). Self-reported compliance with the diet, compliance with the exercise prescription, finding the diet easy to follow, family satisfaction with the changes, satisfaction with energy levels while on the diet, and enjoyment of a variety of foods were not associated with maternal characteristics.

Smartphone App Acceptability

Acceptability of most aspects of the app did not differ by maternal characteristics. Questions that were answered differently by subgroups of women are detailed below. The group who agreed that the graphics on the app were helpful

scored lower for affluence on the HP deprivation index than those who disagreed (4.52 [SD 11.3] vs 10.29 [SD 10.8], respectively; *P*=.03). Those from disadvantaged neighborhoods were more likely to find the graphics helpful than those in advantaged neighborhoods (35/37, 95% [SD 23%] vs 62/81, 77% [SD 43%], respectively; *P*=.03). Those who found the "Tip of the Day" useful had a lower HP index than those who did not (4.55 [SD 11.56] vs 10.33 [SD 7.97], respectively; *P*=.01). Those who agreed that the tip was helpful in planning meals had a lower HP index (2.87 [SD 11.43] vs 9.23 [SD 9.92], respectively; *P*=.002) and were more likely to select this response if from a disadvantaged neighborhood compared to those from an advantaged neighborhood (29/37, 78% [SD 42%] vs 41/81, 51% [SD 50%], respectively; *P*=.008). The group who found the "Exercise of the Day" practical had a lower HP index than those who disagreed (3.83 [SD 12.01] vs 10.11 [SD 7.8], respectively; *P*=.001) and those who were from disadvantaged neighborhoods were more likely to agree than those from advantaged groups (32/37, 86% [SD 35%] vs 52/82, 63% [SD 49%], respectively; *P*=.02). The group who agreed that the "Exercise of the Day" was motivating had a higher median BMI than the group that disagreed (28.74 [IQR 26.7-31.49] vs 27.49 [IQR 26.33-29.22], respectively; *P*=.04). Those who found the "Meal of the Day" useful had a higher median BMI (28.71 [IQR 26.71-31.39) vs 26.74 [IQR 26.18-29.31], respectively; *P*=.02) than those who disagreed. Those from a disadvantaged neighborhood were more likely to agree than those from advantaged neighborhoods (36/39, 92% [SD 27%] vs 61/83, 74% [SD 44%], respectively; *P*=.03). Participants with an obese BMI were more likely to agree that the "Meal of the Day" was motivating than those with an overweight BMI (29/32, 91% [SD 30%] vs 58/87, 67% [SD 47%], respectively; *P*=.02). The group who agreed that the meals on the app were affordable had a lower HP index (4.28 [SD 11.8] vs 8.94 [SD 8.88], respectively; *P*=.02). The group who thought there was enough detail had lower HP index than those who thought there was too little detail (3.87 [SD 10.83] vs 10.32 [SD 11.84], respectively; *P*=.008). Those who achieved less than third-level education were more likely to think there was enough information provided on the app (39/44, 89% [SD 32%] vs 43/62, 69% [SD 47%], respectively; *P*=.04). Younger women were more likely to recommend this app to a friend (32.51 [SD 4.49] vs 35.77 [SD 2.64], respectively; *P*=.03) as well as those with a lower HP index (4.77 [SD 11.18] vs 12.5 [SD 10.93], respectively; *P*=.04). App use was significantly associated with enjoyment of the app, finding the graphics helpful, and acceptability of the "Tip of the day," "Meal of the Day," and "Exercise of the Day" sections of the app (Table 3).

Table 3. Comparison of the objectively measured total app usage instances according to app acceptability by using Mann-Whitney *U* tests.

Question	Agree[a], median (IQR)	Disagree[b], median (IQR)	*P* value
The PEARS[c] app was enjoyable to use (n=114)	28.5 (12-80.75)	11 (6-25)	.006
The PEARS app was easy to use (n=113)	24 (8-73)	10 (4-[d])	.31
The PEARS app was straightforward to follow (n=113)	25 (8-73)	10 (4- [d])	.32
I found the graphics helpful (n=112)	28 (8.5-73)	14 (5.25-42.5)	.04
The PEARS app was useful (n=116)	24 (9.25-72.5)	12 (4.5-68.25)	.28
I found the “Tip of the Day” function useful (n=114)	21 (8-69.5)	28 (6.5-82)	.88
I found the “Tip of the Day” function practical (n=114)	24 (7.5-69.5)	20 (8.5-68)	.88
I found the “Tip of the Day” function motivating (n=114)	26 (8.5-78.25)	17.5 (7-48.75)	.23
I found the “Tip of the Day” function motivated me to eat well (n=112)	28.5 (10.5-83)	15 (6.25-43.75)	.02
I found the “Tip of the Day” function motivated me to be active (n=112)	28.5 (10-84.25)	19 (6.75-41.5)	.06
I found the “Tip of the Day” function was helpful in planning meals (n=113)	24 (10-84.25)	20.5 (7.75-73.5)	.97
I found the “Exercise of the Day” function useful (n=115)	25 (7.25-80)	16 (8-56)	.26
I found the “Exercise of the Day” function practical (n=113)	28 (12-83)	15.5 (6-38)	.02
I found the “Exercise of the Day” function motivating (n=114)	28 (11-76.5)	19 (6-52.5)	.14
I found the “Meal of the Day” function useful (n=116)	25 (7.25-78.25)	20 (8-48.75)	.52
I found the “Meal of the Day” function practical (n=112)	28 (10-81.5)	17 (6-28)	.03
I found the “Meal of the Day” function motivating (n=113)	28 (10-80)	17.5 (6-48.25)	.12
The meals on the app were no more expensive than meals I made before I began the study (n=115)	26 (12-74)	19.5 (6-60.25)	.23
There was enough detail provided on the app (n=116)	24 (10.5-70.25)	23.5 (6.25-85.5)	.74
Would you recommend this app to a friend? (n=117)	24 (8-73)	21.5 (5.5-60.75)	.44
Would you use this app if you were pregnant again? (n=117)	25 (8.75-73)	10 (4-41)	.11

[a]Agree: the sum of the responses “agree” and “strongly agree.”

[b]Disagree: the sum of the responses “disagree” and “strongly disagree.”

[c]PEARS: pregnancy, exercise, and nutrition research study with smartphone app support.

[d]75th percentile was not available in the analysis output.

Multivariate Analysis

In terms of intervention acceptability, in multivariate analysis, enjoyment of the dietary changes undertaken as part of the intervention was significantly associated with a less than third-level education, independent of the effect of confounding factors (Table 4). When further analyzing the app acceptability responses, disadvantaged neighborhood deprivation was significantly associated with agreeing that there is enough detail provided on the app (*P*=.009). There was also a significantly positive association between finding the “Exercise of the Day” motivating and BMI, independent of the effects of confounding factors (*P*=.048) (Table 5). App usage was significantly higher among those who found that the “Tip of the Day” motivated them to eat well (*P*=.03) and who found the “Exercise of the Day” practical (*P*=.004) (Table 6).

Table 4. Multivariate analysis of the association between intervention acceptability and maternal characteristics.[a]

Question	Variable	B[b] coefficient	*P* value	Odds ratio (95% CI)	*P* value[c]
It was easy to follow the diet recommended during the study (n=136)	Pobal Haase-Pratschke index	–0.039	.04	0.962 (0.926- 0.999)	.14
I enjoyed the dietary changes I made (n=137)	Education (achieved third-level)	–1.386	.007	0.25 (0.092-0.683)	.04
The changes I made did not increase my weekly grocery bill (n=136)	Age (years)	0.125	.02	1.133 (1.017-1.262)	.06

[a]Controlling for BMI, age, education level, Pobal Haase-Pratschke index, and parity in a logistic regression model. Only variables that were significant in their respective bivariate models are shown.

[b]Coefficient for the constant in the null model.

[c]Refers to the *P* values in the omnibus test of model coefficients.

Table 5. Multivariate analysis of the association between app acceptability and maternal characteristics.[a]

Question	Variable	B[b] coefficient	*P* value	Odds ratio (95% CI)	*P* value[c]
I found the graphics helpful (n=99)	Pobal Haase-Pratschke index	–0.082	.01	0.922 (0.865-0.982)	.08
I think the app needs more pictures (n=98)	Age	0.087	.11	1.091 (0.98-1.214)	.48
The PEARS[d] app is as good, if not better, than other apps available for pregnancy (n=100)					.05
	Age	–0.115	.04	0.891 (0.801-0.992)	
	Education (achieved third-level)	0.579	.22	1.785 (0.702-4.538)	
I found the “Tip of the Day” function useful (n=101)	Pobal Haase-Pratschke index	–0.046	.15	0.955 (0.896-1.017)	.28
I found the “Tip of the Day” function was helpful in planning meals (n=99)	Pobal Haase-Pratschke index	–0.062	.009	0.94 (0.898-0.984)	.06
I found the “Exercise of the Day” function useful (n=102)					.16
	Pobal Haase-Pratschke index	–0.047	.06	0.954 (0.91-1.001)	
	BMI	0.145	.11	1.156 (0.969-1.38)	
I found the “Exercise of the Day” function practical (n=100)	Pobal Haase-Pratschke index	–0.050	.04	0.951 (0.907-0.998)	.08
I found the “Exercise of the Day” function motivating (n=101)	BMI	0.172	.048	1.188 (1.001-1.409)	.03
I found the “Meal of the Day” function useful (n=103)					.06
	Pobal Haase-Pratschke index	–0.046	.10	0.955 (0.904-1.008)	
	BMI	0.224	.053	1.251 (0.997-1.57)	
I found the “Meal of the Day” function practical (n=98)	Smoking	–22.053	>.99	0 (0)	.09
I found the “Meal of the Day” function motivating (n=100)	BMI	0.204	.02	1.226 (1.029-1.461)	.16
The meals on the app were helpful for preparing breakfast (n=101)	Age	–0.108	.07	0.898 (0.799-1.009)	.41
The meals on the app were no more expensive than meals I made before I began the study (n=102)	Pobal Haase-Pratschke index	–0.050	.04	0.951 (0.906-0.998)	.08
There was enough detail provided on the app (n=103)					.01
	Pobal Haase-Pratschke index	–0.082	.009	0.921 (0.866-0.979)	
	Education (achieved third level)	0.762	.24	2.142 (0.601-7.63)	
Would you recommend this app to a friend? (n=104)					.15
	Age	–0.228	.04	0.796 (0.642-0.988)	
	Pobal Haase-Pratschke index	–0.048	.23	0.953 (0.88-1.032)	

[a]Controlling for BMI, age, education level, Pobal Haase-Pratschke index, and parity in a logistic regression model. Only variables that were significant in their respective bivariate models are displayed above.

[b]Coefficient for the constant in the null model.

[c]Refers to the *P* values in the omnibus test of model coefficients.

[d]PEARS: pregnancy, exercise, and nutrition research study with smartphone app support.

Table 6. Association between app acceptability and app instances of use (obtained from the app's software).[a]

Question	B[b] coefficient	*P* value	Odds ratio (95% CI)	*P* value[c]
The PEARS[d] app was enjoyable to use (n=114)	0.026	.04	1.027 (1.002-1.052)	.10
I found the graphics helpful (n=112)	0.018	.07	1.018 (0.998-1.038)	.04
I found the "Tip of the Day" function motivated me to eat well (n=112)	0.021	.02	1.021 (1.003-1.039)	.03
I found the "Exercise of the Day" function practical (n=113)	0.020	.009	1.02 (1.005-1.035)	.004
I found the "Meal of the Day" function practical (n=112)	0.011	.13	1.011 (0.997-1.025)	.06
I found the "Meal of the Day" function appetizing (n=113)	0.012	.07	1.012 (0.999-1.025)	.42
The meals on the app were helpful for preparing dinner (n=116)	0.007	.34	1.007 (0.993-1.02)	.37

[a]Controlling for smoking, age, Pobal Haase-Pratschke index, education level, BMI, parity, and ethnicity in a logistic regression model. Only variables that were significant in their respective bivariate models are displayed above.

[b]Coefficient for the constant in the null model.

[c]Refers to the *P* values in the omnibus test of model coefficients.

[d]PEARS: pregnancy, exercise, and nutrition research study with smartphone app support.

Qualitative Interviews

Interviews were conducted with 28 PEARS participants whose characteristics are outlined in Table S1 of Multimedia Appendix 4. The majority of the participants were very satisfied with the content of the PEARS study.

> *…Like truthfully, it all worked for me. Like I had the app to give me ideas and I had you and (the PEARS study doctor) if I needed supports. So for me, there really isn't anything I would change. And I'm not saying that to be nice. I honestly, it was fine, it was helpful for me and it motivated me. But it never took over, and that's what I liked about it.* [Sheila, intervention, 39 years old, nulliparous, BMI 28, White Irish, second-level education]

Diet Intervention

Most women made positive changes to their diet since participation in the PEARS study. The majority of women found the dietary changes easy to make and reported on average that they complied with the recommendations approximately 80% of the time.

> *…No it wasn't actually that difficult, when I got into the swing of it. Now I'd say the first couple of weeks, I was very conscious of what I was doing. And then it was sure nothing.* [Nicole, intervention, 36 years old, parity 4, BMI 25, White Irish, third-level education]

The wide variety of food options when following a low-GI diet and the idea of swapping foods rather than eliminating them made the recommendations more acceptable and sustainable for participants. However, there were varying levels of acceptability of the financial cost of following a low-GI diet. Some participants reported that healthy foods were more expensive but that this was balanced by the avoidance of certain snack foods.

> *…Not really I was buying a lot more fruit, so that was a bit more expensive and the kind of salad stuff and the veg kind of would bring the shopping bill up a bit, but I mean I think it was worth it like and as I said I cut out buying crisps and chocolate covered biscuits and stuff so then it was coming down as well.* [Orla, intervention, 30 years old, parity 1, BMI 25, White Irish, third-level education]

Exercise Intervention

There was a substantially less positive change in physical activity behavior during pregnancy. Time constraints were cited as the most common barrier to exercise.

> *…If I had time, [if] it was my first pregnancy, I'd probably have a lot more time, more effort.* [Lianne, intervention, 37 years old, parity 1, BMI 25, White Irish, second-level education]

However, the encouragement of physical activity made women with low self-efficacy perceive physical activity as accessible.

> *…Even with the exercise say, with the 30 minutes a day, when you told me oh you can split that up over three 10 minutes sure fabulous, like who can't fit in three 10 minutes a day?* [Sally, intervention, 31 years old, nulliparous, BMI 26, Irish, third-level education]

Smartphone App Intervention

Overall, the women appeared very satisfied with the app and the majority agreed that it played a role in their adherence to the PEARS dietary recommendations. Participants valued the reliability of the PEARS app as the content was created and approved by a multidisciplinary team.

> *…I know with your app that's 100 per cent guaranteed - it has the backup, it has all the right information.* [Caoimhe, intervention, 34 years old, parity 3, BMI 34, White Irish, second-level education]

The "Tip of the Day" was the strongest enticement for interview participants to access the app.

> *…I probably would go into the app purposely on a daily basis to read the 'Tip of the Day'.* [Laura,

intervention, 26 years old, nulliparous, BMI 28, Irish, third-level education]

Some interview participants reported a decline in their use of their app throughout the intervention. This was particularly evident among women who had achieved tertiary education due to perceived confidence in their cooking abilities and knowledge.

> *…I think it's because I thought I was familiar with the material. I think it's because our meeting kind of armed me and I felt like okay well I know what I'm not supposed to do and I know what I can do, so just do that and then I didn't check in with it you know. … I thought that the meals I was having already would probably conform to the diet.* [Kelly, intervention, 34 years old, nulliparous, BMI 26, White Irish, third-level education]

Discussion

Principal Results

The PEARS diet and lifestyle intervention with smartphone app support was accepted by a pregnant cohort who were overweight or obese and greater acceptability was associated with lower education level. The use of a smartphone app in the intervention was deemed highly acceptable and aspects of acceptability were associated with lower socioeconomic status and higher BMI. Enjoyment of certain sections of the app was associated with increased use, as measured by the app's software.

Mixed Lifestyle Intervention Acceptability

The acceptability of mixed lifestyle interventions has been reported in previous studies among similar cohorts of pregnant women [4,7,19]. Participants of PEARS reported to be compliant with the dietary aspect of the intervention and found a low-GI diet enjoyable and easy to follow. In a previous study of a low-GI diet in pregnancy (ROLO Study), which did not incorporate an exercise or mHealth component, acceptability of the diet was also high [20]. Similar results were found in both the studies; in the ROLO study, 68% found the low-GI diet easy to follow and 65% found it enjoyable, while in the PEARS study, 69% found it easy to follow and 74% found it enjoyable [20]. The use of smartphone app technology as part of the PEARS intervention was deemed acceptable by the cohort of pregnant women who were overweight or obese, as reported in both the quantitative and qualitative analyses. High acceptability of the use of mHealth was also reported in other mixed lifestyle interventions among pregnant women with a high BMI [4,6,7] and highlights mHealth as a positive way for health care professionals to communicate reliable health and lifestyle information to their patients. Having a higher BMI was associated with acceptability of the "Exercise of the Day" subsection of the app. A study among college-aged women has shown that high BMI is associated with experiences of weight stigma, which is associated with exercise avoidance [25]. As those with high BMI particularly enjoyed the exercise portion of the PEARS app, the app may have been successful in overcoming this barrier and encouraging the enjoyment of exercise.

Engagement With Women of Low Socioeconomic Status

Low education level is often associated with poor dietary habits and poor response to dietary interventions [26,27] and has been associated with gestational weight gain outside the recommendations in pregnancy [28]. Contrastingly, this study reports higher acceptability of a dietary intervention among participants with lower education level, a group that may have been required to make the greatest change to their preintervention diet to comply with the recommended diet. The PEARS dietary intervention is accepted by a group with higher likelihood of having poor dietary habits before pregnancy and who are more likely to exceed the recommended gestational weight gain [26-28]. Therefore, this group may have the most potential to benefit from a dietary intervention. The qualitative analysis suggests that women with higher educational attainment more commonly discontinued their use of the app throughout the course of the intervention, as they felt confident in their knowledge and cooking abilities. A smartphone app with recipes and meal ideas may be a useful tool for those with less confidence in their cooking abilities and nutrition knowledge. This study-specific smartphone app was deemed easy to use and to understand by all participants, including those from hard-to-reach subgroups of the population. Low socioeconomic status was associated with acceptability of the level of detail provided on the app. The addition of mHealth technology to the PEARS study enabled women with all levels of literacy and previous nutrition knowledge to engage with the intervention. If used in future studies or for public health purposes, smartphone apps may be adapted to the literacy levels of the target population to assist in improving health outcomes. Our findings support mHealth apps as a tool that may help in bridging the socioeconomic health gap, which exists in Ireland, by providing an adequate and accessible source of additional health care information [9].

Clinical Implications

The "Tip of the Day" function was reported to be the most useful subsection of the app, as reported in the app questionnaire. Results from the qualitative interviews support this finding, as participants reported that this feature was their highest motivator in accessing the app. This section provided a small piece of information in a concise manner. This finding is highly relevant to the current health care climate in Ireland, which is introducing the concept of "Making Every Contact Count" (MECC). It is a health behavior change framework that will be implemented to take advantage of the interactions health care professionals have with patients in order to send positive health care messages, which ultimately aim to reduce the incidence of chronic disease [28]. The basic level of its implementation is giving patients brief healthy lifestyle advice when the opportunity arises. The MECC framework plans to incorporate mHealth as a tool to communicate with both patients and staff. A smartphone app that contains a feature such as "Tip of the Day," which delivers brief diet and exercise advice may be of use in such a behavior change framework, as in this study, it was found to be acceptable in a pregnant cohort at a higher risk of developing GDM. mHealth also has the potential to support current dietetic practice

as part of individual patient counselling or to deliver novel and convenient nutrition information to hard-to-reach patient groups.

Future of Health Care Apps

Most participants stated they would use the app in pregnancy again and recommend it to others. Of the entire PEARS cohort of 565 participants, 379 (67.1%) participants in early pregnancy reported using a pregnancy app as a source of information (unpublished data from the PEARS study). Although this figure may encompass a wide variety of pregnancy apps (eg, gestation calculators, pregnancy tracking, health information), it is known that there is a lack of regulation across all types of commercial pregnancy apps [11]. This highlights the demand for the development of suitable apps, designed and approved by health care professionals, particularly pregnancy apps that provide diet and lifestyle advice. Access to reliable diets and lifestyle information in a pregnancy app also appears to be important to pregnant women, as interview participants valued that the PEARS app contained reliable information, approved by health care professionals. App usage was higher among those who found the "Tip of the Day" motivating and those who found the "Exercise of the Day" practical. This demonstrates that the delivery of information in a concise format and practical information and advice were valuable to participants. Features that make the app clear and non–time-consuming would be important aspects to consider if creating future mHealth apps for similar populations.

Strengths and Limitations

The strengths of this study include the large population size and the availability of app usage data in order to verify participant app use. The cohort studied is well-characterized and a large amount of data was collected on maternal characteristics in early pregnancy. This was also a representative sample of the PEARS study (Multimedia Appendix 3). This is the first pregnancy study to report on the acceptability of an app used to deliver healthy lifestyle information, thus adding to this field. The availability of both quantitative and qualitative data was also a strength of this research.

A limitation of this study was that compliance and acceptability of both the intervention and the app were self-reported, thereby possibly introducing incorrect reporting. Compliance with recommended app use could be verified using the app's software; however, diet and exercise compliance were not verified by any other measurements. Another limitation was that the questions in the questionnaires were not open-ended. Participants were limited to choosing from a list of responses that were preselected by the research team, which may have influenced their true opinion. However, this is also a strength as it allows for statistical analysis. These findings are specific to a pregnant cohort with a BMI ≥25 kg/m², which is not representative of the BMI of a general pregnant population. However, given that half of pregnant women in Ireland and other Western populations have a BMI in the overweight or obese categories [29,30], our findings are relevant for all clinicians working in the antenatal setting. As individuals who volunteer for research studies typically have higher education and are older than those who decline participation [31], the generalizability of our findings to a broader population group may be affected. Responses were also not received from all participants of the intervention group for unknown reasons; however, there were no differences in the maternal demographics between respondents and nonrespondents. An important limitation to this research was 174 (9.4%) women who were assessed for eligibility for the PEARS study (n=1858) did not own a smartphone and were therefore ineligible to participate in the intervention. Although results have demonstrated that aspects of the app showed higher acceptability among participants from lower socioeconomic backgrounds, the exclusion criteria of the study may have excluded the most hard-to-reach women, who may not be in possession of a smartphone. It is also evident from the app usage data that the app was not used as often as recommended by researchers. Results demonstrate that app usage was associated with acceptability of certain aspects of the app; usage was higher among participants who enjoyed "Exercise of the Day" and "Tip of the Day." Therefore, if the app had been used more regularly, perhaps the participants' acceptability may have differed.

Conclusions

A low-GI diet and physical activity intervention with smartphone app support was accepted by a population of pregnant women who were overweight or obese. The diet was considered enjoyable and easy to follow, and compliance was high. The acceptability of the smartphone app was also high. Although the smartphone app used in the PEARS study was not targeted at a population of specific socioeconomic status or education level, results demonstrate that aspects of the app were particularly accepted by these hard-to-reach groups. The app appealed to those who had a higher BMI and were therefore more at risk of developing GDM. When developing a future app for a similar population, the socioeconomic and educational characteristics of the population should be considered in order to adapt the level of detail and presentation of information accordingly.

Acknowledgments

The authors thank the women of the PEARS study for sharing their time and data and Dr Ricardo Segurado for statistical support. The PEARS randomized controlled trial was funded by University College Dublin and The National Maternity Hospital Medical Fund. The funding source did not play any role in the trial design or manuscript presentation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention acceptability questionnaire.
[PDF File (Adobe PDF File), 225 KB - mhealth_v9i5e17189_app1.pdf]

Multimedia Appendix 2
App acceptability questionnaire.
[PDF File (Adobe PDF File), 321 KB - mhealth_v9i5e17189_app2.pdf]

Multimedia Appendix 3
Comparison of maternal demographics between questionnaire respondents and nonrespondents.
[PDF File (Adobe PDF File), 156 KB - mhealth_v9i5e17189_app3.pdf]

Multimedia Appendix 4
Characteristics of the respondents to the qualitative interviews.
[PDF File (Adobe PDF File), 83 KB - mhealth_v9i5e17189_app4.pdf]

References

1. mHealth: new horizons for health through mobile technologies. World Health Organisation Global Observatory for eHealth. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-04-14]
2. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
3. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to Health: A Novel mHealth Intervention Using Fitbits and Twitter to Foster Healthy Lifestyles. Clin Pediatr (Phila) 2017 Jan;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
4. Soltani H, Duxbury A, Arden M, Dearden A, Furness P, Garland C. Maternal Obesity Management Using Mobile Technology: A Feasibility Study to Evaluate a Text Messaging Based Complex Intervention during Pregnancy. J Obes 2015;2015:814830 [FREE Full text] [doi: 10.1155/2015/814830] [Medline: 25960889]
5. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: A randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
6. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014 Nov;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
7. Willcox J, Wilkinson S, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
8. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women - the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017 Mar 23;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]
9. IMO position paper on health inequalities. Irish Medical Organisation. URL: https://www.imo.ie/policy-international-affair/documents/imo-position-papers/imo-position-paper-on-hea/IMOPP-Health-Inequalities-(2)-.pdf [accessed 2021-04-08]
10. Healthy Ireland survey 2017 summary of findings. Healthy Ireland. URL: https://assets.gov.ie/9542/8aa3c0d3f1d8417c95b025b69bc0c8d3.pdf [accessed 2021-04-08]
11. Hughson J, Daly J, Woodward-Kron R, Hajek J, Story D. The Rise of Pregnancy Apps and the Implications for Culturally and Linguistically Diverse Women: Narrative Review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
12. Kennelly M, Ainscough K, Lindsay K, O'Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy Exercise and Nutrition With Smartphone Application Support: A Randomized Controlled Trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]
13. Lee AJ, Hiscock RJ, Wein P, Walker SP, Permezel M. Gestational diabetes mellitus: clinical predictors and long-term risk of developing type 2 diabetes: a retrospective cohort study using survival analysis. Diabetes Care 2007 Apr;30(4):878-883. [doi: 10.2337/dc06-1816] [Medline: 17392549]
14. Coustan D. Gestational diabetes mellitus. Clin Chem 2013 Sep;59(9):1310-1321. [doi: 10.1373/clinchem.2013.203331] [Medline: 23536513]
15. Torloni M, Betrán AP, Horta B, Nakamura MU, Atallah AN, Moron AF, et al. Prepregnancy BMI and the risk of gestational diabetes: a systematic review of the literature with meta-analysis. Obes Rev 2009 Mar;10(2):194-203. [doi: 10.1111/j.1467-789X.2008.00541.x] [Medline: 19055539]

16. Walsh JM, McGowan CA, Mahony R, Foley ME, McAuliffe FM. Low glycaemic index diet in pregnancy to prevent macrosomia (ROLO study): randomised control trial. BMJ 2012 Aug 30;345:e5605 [FREE Full text] [doi: 10.1136/bmj.e5605] [Medline: 22936795]
17. Wang C, Wei Y, Zhang X, Zhang Y, Xu Q, Sun Y, et al. A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women. Am J Obstet Gynecol 2017 Apr;216(4):340-351. [doi: 10.1016/j.ajog.2017.01.037] [Medline: 28161306]
18. Brown J, Alwan N, West J, Brown S, McKinlay CJ, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017 May 04;5:CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
19. Warren L, Rance J, Hunter B. Feasibility and acceptability of a midwife-led intervention programme called 'Eat Well Keep Active' to encourage a healthy lifestyle in pregnancy. BMC Pregnancy Childbirth 2012 Apr 11;12:27 [FREE Full text] [doi: 10.1186/1471-2393-12-27] [Medline: 22494540]
20. McGowan CA, Walsh JM, Byrne J, Curran S, McAuliffe FM. The influence of a low glycemic index dietary intervention on maternal dietary intake, glycemic index and gestational weight gain during pregnancy: a randomized controlled trial. Nutr J 2013 Oct 31;12(1):140 [FREE Full text] [doi: 10.1186/1475-2891-12-140] [Medline: 24175958]
21. Kennelly MA, Ainscough K, Lindsay K, Gibney E, Mc Carthy M, McAuliffe FM. Pregnancy, exercise and nutrition research study with smart phone app support (Pears): Study protocol of a randomized controlled trial. Contemp Clin Trials 2016 Jan;46:92-99. [doi: 10.1016/j.cct.2015.11.018] [Medline: 26625980]
22. Haase T, Pratschke J. The 2016 Pobal HP deprivation index for small areas (SA): Introduction and reference tables. URL: https://www.pobal.ie/app/uploads/2018/06/The-2016-Pobal-HP-Deprivation-Index-Introduction-07.pdf [accessed 2021-04-08]
23. Moses R, Luebcke MFDW, Davis WFCK, Coleman KJ, Tapsell LC, Petocz P, et al. Effect of a low-glycemic-index diet during pregnancy on obstetric outcomes. Am J Clin Nutr 2006 Oct;84(4):807-812. [doi: 10.1093/ajcn/84.4.807] [Medline: 17023707]
24. Frederick CM, Ryan RM. Differences in motivation for sport and exercise and their relations with participation and mental health. Journal of Sport Behavior. 1993. URL: https://www.researchgate.net/publication/313702068_Differences_in_Motivation_for_Sport_and_Exercise_and_their_Relations_with_Participation_and_Mental_Health [accessed 2021-04-14]
25. Vartanian LR, Shaprow JG. Effects of weight stigma on exercise motivation and behavior: a preliminary investigation among college-aged females. J Health Psychol 2008 Jan;13(1):131-138. [doi: 10.1177/1359105307084318] [Medline: 18086724]
26. Friis K, Lasgaard M, Rowlands G, Osborne RH, Maindal HT. Health Literacy Mediates the Relationship Between Educational Attainment and Health Behavior: A Danish Population-Based Study. J Health Commun 2016;21(sup2):54-60. [doi: 10.1080/10810730.2016.1201175] [Medline: 27668691]
27. O'Brien EC, Alberdi G, Geraghty AA, McAuliffe FM. Lower education predicts poor response to dietary intervention in pregnancy, regardless of neighbourhood affluence: secondary analysis from the ROLO randomised control trial. Public Health Nutr 2017 Aug 15;20(16):2959-2969. [doi: 10.1017/s1368980017001951]
28. O'Brien E, Alberdi G, McAuliffe F. The influence of socioeconomic status on gestational weight gain: a systematic review. J Public Health (Oxf) 2018 Mar 01;40(1):41-55. [doi: 10.1093/pubmed/fdx038] [Medline: 28398550]
29. Mahony R. Clinical Report for the Year 2016. Dublin: The National Maternity Hospital; 2016.
30. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA 2012 Mar 01;307(5):491-497. [doi: 10.1001/jama.2012.39] [Medline: 22253363]
31. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]

Abbreviations

GDM: gestational diabetes mellitus
GI: glycemic index
HP: Haase-Pratschke
MECC: Making Every Contact Count
mHealth: mobile health
PEARS: pregnancy, exercise, and nutrition research study with smartphone app support

Edited by G Eysenbach; submitted 25.11.19; peer-reviewed by M Bijlholt, KH Tan; comments to author 21.12.19; revised version received 10.04.20; accepted 14.05.20; published 12.05.21.

Please cite as:
Greene EM, O'Brien EC, Kennelly MA, O'Brien OA, Lindsay KL, McAuliffe FM
Acceptability of the Pregnancy, Exercise, and Nutrition Research Study With Smartphone App Support (PEARS) and the Use of Mobile Health in a Mixed Lifestyle Intervention by Pregnant Obese and Overweight Women: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e17189
URL: https://mhealth.jmir.org/2021/5/e17189
doi:10.2196/17189
PMID:33978597

Original Paper

Effects of Psychoeducational Interventions Using Mobile Apps and Mobile-Based Online Group Discussions on Anxiety and Self-Esteem in Women With Breast Cancer: Randomized Controlled Trial

Elaheh Ghanbari[1], BSc, MSc; Shahrzad Yektatalab[2], BSc, MSc, PhD; Manoosh Mehrabi[3], BSc, MSc, PhD

[1]Department of Mental Health and Psychiatric Nursing, School of Nursing and Midwifery, Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran

[2]Department of Mental Health and Psychiatric Nursing, School of Nursing and Midwifery, Community-Based Psychiatric Care Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

[3]Department of E-Learning Planning in Medical Sciences, Virtual School, Shiraz University of Medical Sciences, Shiraz, Iran

Corresponding Author:
Shahrzad Yektatalab, BSc, MSc, PhD
Department of Mental Health and Psychiatric Nursing
School of Nursing and Midwifery, Community-Based Psychiatric Care Research Center
Shiraz University of Medical Sciences
Zand Street, Namazee Square
Shiraz
Iran
Phone: 98 917 710 6721
Fax: 98 713 647 4250
Email: shahrzadyekta@yahoo.com

Abstract

Background: Psychoeducation has turned into an effective tool in taking care of cancer patients and improving their psychophysical symptoms and quality of life. Despite the growing use of mobile phone apps in medical settings for improving health, evidence supporting their effectiveness in the psychoeducation of patients with breast cancer is rarely available.

Objective: This study was conducted to investigate the effect of psychoeducational interventions on anxiety and self-esteem in women with breast cancer using a mobile app and an online support group.

Methods: An unblinded randomized controlled trial based on mobile phones was conducted in Shiraz, Iran. A research assistant recruited 82 women with nonmetastatic breast cancer aged 20 to 60 years were from clinics during a face-to-face visit at the point of care and randomly assigned to an intervention group (n=41) and a wait-list control group (n=41) through blocked randomization. The intervention group received psychoeducational interventions through a mobile phone app and participated in nurse-assisted online mobile support sessions for a total four weeks, whereas the control group was put on a waiting list. The State-Trait Anxiety Inventory (STAI) and the Rosenberg Self-Esteem Scale (RSES) were used to measure the levels of anxiety and self-esteem as the main outcomes at baseline and one week after the intervention.

Results: A total of 82 patients with a mean age of 46.45 (SD 9.29) years recruited in Winter 2016 were randomly assigned to a wait-list control group (n=41) and intervention group (n=41). Five patients dropped out for different reasons. Comparing the postintervention mean scores of anxiety and its subscales using the independent t test showed statistically significant differences between the mobile psychoeducation group and controls (P<.001). The paired t test used to compare the postintervention mean scores of anxiety with its preintervention scores in the intervention group showed significant reductions in the scores of anxiety (95% CI –17.44 to –8.90, P<.001, d=1.02) and its two subscales (state anxiety: 95% CI –9.20 to –4.21, P<.001, d=0.88 and trait anxiety: 95% CI –8.50 to –4.12, P<.001, d=0.94). Comparing the postintervention mean scores of self-esteem showed statistically insignificant differences between the control and intervention groups (16.87 vs 17.97, P=.24). In contrast with the controls, using the paired t test showed that the increase in the postintervention mean scores of self-esteem were statistically significant in the intervention group compared with the preintervention scores (mean difference 2.05, 95% CI 1.28 to 2.82, P<.001).

Conclusions: This study demonstrated the key role of mobile apps in decreasing anxiety and improving self-esteem in women with breast cancer through psychoeducational interventions. Similar studies with longer follow-ups are recommended that be conducted in this context.

Trial Registration: Iranian Registry of Clinical Trials IRCT2015072123279N2; https://en.irct.ir/trial/19882

(JMIR Mhealth Uhealth 2021;9(5):e19262) doi:10.2196/19262

KEYWORDS

anxiety; breast cancer; education; mobile app; self-esteem

Introduction

As the most prevalent cancer in women and the second most prevalent cancer, breast cancer accounts for 25.2% of cancers globally and 24.6% of cancers in Iran [1,2]. Advanced sciences, early diagnosis, and timely treatments have increased the survival rate of patients with breast cancer. The long-term survival of patients with breast cancer, however, predisposes them to many psychological disorders [3].

Psychological distress is highly prevalent among women with breast cancer, and they are at increased risks for developing anxiety. The prevalence of anxiety was found to be 24.1% in patients with breast cancer [4]. Research suggests higher levels of anxiety are associated with increased fear of recurrence, hopelessness, uncertainty, and loss of control and lower levels of satisfaction with life. Cancer-associated anxiety can intensify the disease symptoms in patients, prolong their recovery, cause undesirable outcomes, and degrade their and their caregivers' quality of life [5,6]. Anxiety is the most common response to a mastectomy, potentially in association with body image, pain, and disfigurement, and is reported to be normally associated with a sense of disability that lowers self-esteem in women who have undergone a mastectomy [7,8].

Moreover, the changes caused by breast cancer and its treatment, mastectomy, changes in roles and functions, fear of losing femininity, sexual dysfunction, and impaired body image are considered major factors that lower self-esteem in patients [9]. Research has demonstrated that self-esteem was found to be adversely affected in women with breast cancer, which in turn can further impair their psychological well-being [10]. Moreover, low self-esteem was found to be associated with anxiety in this group of women [11].

High risks for developing anxiety and low self-esteem have been repeatedly reported in women with breast cancer. These patients therefore require special support in different stages, including diagnosis, treatment, rehabilitation, and follow-up [8,12].

Given the overwhelming effects of anxiety and low self-esteem on mental health and quality of life in women with breast cancer and their families and caregivers, effective psychological interventions (eg, psychoeducation and cognitive behavioral therapy) are recommended to be designed to improve quality of life and functions in these patients, shorten their recovery, and improve their prognosis [13,14]. Research suggests beneficial effects of psychoeducational interventions on patients with breast cancer [15,16].

Clinicians and patients have increasingly used mobile devices in the last decade. Given the growing use of mobile health apps and devices in medical sciences and the health care industry, the World Health Organization defined mobile health (mHealth) as using mobile and wireless technologies to help achieve health objectives [17]. mHealth apps include patient education, disease self-management, remote monitoring of patients, diagnostic treatment services, data collection, communication, and counseling services [18].

With a key role in the management and delivery of cancer care, mHealth can assist health care professionals and patients with the diagnosis of cancers and their associated psychological distress as well as follow-up, planning, providing cancer-related information, supporting medication adherence, and managing side effects [19].

Bender et al [20] reported hundreds of cancer-focused apps with the potential for conveniently providing real-time support interventions, monitoring a host of symptoms and physiological indicators of the disease, and promoting behavioral changes in a cost-effective manner; nevertheless, there is a lack of evidence for their safety and effectiveness, and little is known about interventions based on smartphones because many studies lacked a comparison or control group. Recently conducted studies support the effectiveness and efficacy of mobile and internet interventions [21,22]. Moreover, few standardized valid apps in the oncological field exist that can help with providing cancer care and supporting patients during their treatment and follow-up [23].

In addition to promoting health-related quality of life and decreasing stress, mHealth-based interventions were found to improve weight loss and physical activity in patients with breast cancer [24-26]; nevertheless, the evidence for the effects of mHealth apps and mental health interventions on psychological dimensions in patients with breast cancer is inconclusive and conflicting. According to Foley et al [27], acquiring information about breast cancer decreases anxiety levels before and after a breast cancer surgery; however, they reported higher anxiety levels after surgery in women who used mHealth apps compared with controls.

Although patients rated self-esteem as extremely high and very important in mHealth intervention, this dimension has not been included in previous psycho-oncological mHealth interventions [28].

Given that systematic reviews of mHealth apps in breast cancer are mainly limited to feasibility or pilot studies, the validity of their results is questionable [29]. To the best of our knowledge,

randomized controlled trials (RCTs) have rarely addressed the effectiveness of mobile psychoeducational apps in self-esteem and anxiety in women with breast cancer given the large number of available mHealth apps. Although the efficacy of app-based interventions in reducing the symptoms of depression and anxiety has been demonstrated in RCTs, mHealth apps rarely take an expert approach, adhere to relevant medical evidence, or reflect patient needs. Research suggests using an app coupled with human support (eg, coaching and supervision through phone or text messages) can help increase engagement and use and promote the outcomes [30,31]; for instance, a mobile phone app supported with a behavioral health intervention based on social media was found to promote quality of life and psychosocial and physiological outcomes in breast cancer survivors within 10 weeks [32].

Given the positive effects of mHealth, this study developed an mHealth phone app known as the Breast Cancer Support zone (BCSzone) to help satisfy the needs of patients with breast cancer for psychoeducational interventions. This study aimed at exploring the effects of psychoeducational interventions on anxiety and self-esteem in women with breast cancer using the mobile app and mobile-based online group discussions. Our study hypothesis states that mobile-based psychoeducational interventions can improve self-esteem and reduce anxiety in patients with breast cancer.

Methods

Study Design

This RCT used a pretest and posttest design to investigate the impact of a mobile app-based psychoeducational intervention on anxiety and self-esteem in women with nonmetastatic breast cancer presenting to breast clinics in Shiraz, Iran. The patients were randomly assigned to the wait-list control group and intervention group receiving the psychoeducational interventions through the mobile phone app.

Setting and Sample

In winter 2016, recruitment was performed in two breast clinics affiliated with Shiraz University of Medical Sciences, Shiraz, Iran, during clinic health appointments. Recruitment was achieved with posters and flyers about the study distributed in clinics, online announcements on the website of Shiraz University of Medical Sciences, and direct referrals by oncologists and health care providers at the study settings.

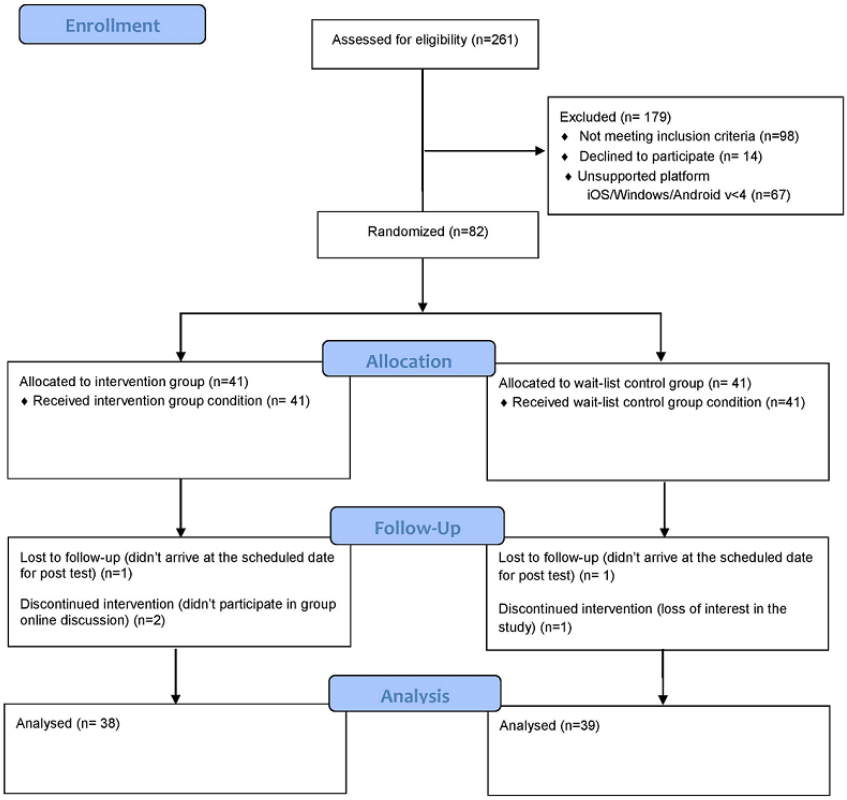
In the recruitment step, patients were briefed on the study objectives and procedure through face-to-face communication and provided with a brochure containing the aim and basic information about the study in the reception of the breast clinics and examination rooms. Clinicians at each clinic were furnished with information regarding the study, and the researchers also responded to the inquiries and concerns of the participants. After informed consent was obtained, participants completed all baseline questionnaires and received a 45-day package of free mobile data (50 GB) for their participation in the study.

The inclusion criteria comprised age range of 20 to 60 years, willingness to participate in the study, diagnosis of nonmetastatic breast cancer, literacy, access to smart mobile electronic devices connected to the internet and willingness to have the app installed on them, ability to work with the app and social networks, and moderate-to-severe anxiety (State-Trait Anxiety Inventory [STAI] score of >80 [33]) and low-to-moderate self-esteem (Rosenberg Self-Esteem Scale [RSES] score of <25 [34]). The exclusion criteria consisted of failing to regularly participate in the educational therapeutic program, serious disease other than breast cancer, history of chronic psychological disorders and taking psychiatric drugs, and participation in similar psychoeducational programs, which could have biased the results.

The minimum sample size was calculated as 25 per group in Number Cruncher Statistical System (NCSS) based on the data obtained by Aghabarari et al [35], mean anxiety difference score of 3 (SD 2.08), effect size of 0.79, significance level of .05, and test power of 80%. The ultimate sample size was calculated as 41 in each group considering the dropout effect.

In winter 2016, out of 261 women with breast cancer who volunteered to participate in the study, 82 were found to be eligible. Random Allocation Software (version 1.0.0, Saghaei M, Isfahan University of Medical Sciences, Iran) was used to generate a randomization schedule. After signing written informed consent forms, the women were randomly assigned to the intervention group (n=41) or the wait-list control (n=41); group randomization was performed by the researcher's assistant. The study flow is illustrated in Figure 1.

Figure 1. Consolidation Standards of Reporting Trials diagram.



As shown in Figure 1, randomization and attrition data were organized according to CONSORT (Consolidated Standards of Reporting Trials) guidelines, with 2 patients in the control group and 3 in the intervention group dropping out due to failure to complete the intervention (n=2), unwillingness to continue with the study (n=1), and lost to follow-up (n=2). The control group was put on a waiting list and only received routine health care, whereas the intervention group received psychoeducational interventions through a mobile app and participated in mobile-based online group discussions for 1 month. Pretest and posttest data were collected using paper-based instruments. The outcome assessors were different from the researchers who conducted and monitored the interventions.

Ethical Considerations

This study was approved by the Ethics Committee of Shiraz University of Medical Sciences, Shiraz, Iran (IR.SUMS.REC.1395.20), and registered in the Iranian Registry of Clinical Trials [IRCT2015072123279N2] [36]. Written informed consent was obtained from all the study participants, and their voluntary participation in the entire process of the study and confidentiality of their information were ensured. All participants were allowed adequate time to carefully review the consent forms and ask relevant questions before signing them. At the end of the study, all patients in the wait-list control group also received the psychoeducational intervention.

Outcome Measures

The data collection tools comprised the STAI, RSES, and demographic information questionnaires, including age, marital status, level of education, employment status, and history of mastectomy. The STAI and RSES were used to measure anxiety and self-esteem as the study variables and compare them in the 2 groups before and 1 week after the intervention. At the end of the treatment, the intervention group completed a satisfaction survey of the mobile-based intervention designed by the authors (Multimedia Appendix 1).

Anxiety

Anxiety was measured with the 40-item STAI as a self-report instrument that quantifies adult anxiety. The STAI comprises two 20-item subscales, state and trait anxiety. The state anxiety subscale evaluates the current anxiety status by determining how the respondent feels right now, whereas the trait anxiety subscale assesses how individuals generally and usually feel. Each item is scored on a 4-point Likert scale, resulting in a total subscale score of 20 to 80, with higher scores denoting higher anxiety levels. The total STAI score therefore equals 40 to 160 [37]. This valid and reliable questionnaire introduced by Spielberger et al [37] in 1970 has been translated to many different languages. An overall median Cronbach alpha of .86 to .92 has also confirmed the internal consistency of this instrument in normative samples [37-39]. A reliability coefficient of .89 to .90 was reported in the literature for STAI-T and .86 to .95 for STAI-S in diverse populations and cultures. A Cronbach alpha of .886 for trait anxiety and .846 for state anxiety also confirmed the internal consistency of the Persian version of the STAI [39,40].

Self-Esteem

The 10-item RSES was used to measure self-esteem on a 4-point Likert scale ranging from strongly agree=0 to strongly disagree=3. The total score is 0 to 30, with higher scores denoting higher self-esteem. Rosenberg reported a reliability

coefficient of .77 and a validity coefficient of .82 for this scale. Administering the RSES in 53 nations, Schmitt et al [41] also confirmed the validity and reliability of this test. Furthermore, the RSES has been translated and adapted to different languages, including Persian. Validity and reliability of the Persian version of the RSES have been confirmed previously [42]. Alizade et al [43] reported a reliability coefficient of .73 and a validity coefficient of .74 in Iran. In another study, the Cronbach alpha coefficient of the Persian version of the scale was calculated as .84 [44]. In this study, Cronbach alpha was reported .70 for the baseline reliability coefficient of the scale.

Survey of satisfaction with the mobile-based intervention

This 16-item author-designed survey assessed the attitude and satisfaction of the participants with the designed mobile app. The item scores were based on respondent selections of very frequently, frequently, occasionally, rarely, and very rarely.

Intervention

Instructional Design

In the first step, BCSzone was developed as an Android-based smartphone app that provided users with educational materials, including texts, images, animations, quizzes, audio files, and video clips, that demonstrated how to accurately perform the exercises. This app included educational topics and psychological and practical exercises and tests to be used offline by the users. Screenshots of the mobile app are available in Multimedia Appendix 2.

The instructional design of this mobile-based intervention was based on a model developed in a virtual school, the Comprehensive Center of Excellence for E-Learning in Medical Sciences, Shiraz University of Medical Sciences. This model was used to design multimedia-based education and comprised the following steps: planning and analysis, text authoring and design, production and preparation of every medium, developing scenarios, ultimately evaluating the educational material and metadata preparation and storage, and technical evaluation and presentation. All these steps focused on the stage evaluation to provide proper feedback for each stage [45,46]. The e-content was developed using Adobe Flash in predesigned formats and based on the standards of the virtual school. After evaluating, editing, and modifying the produced content in alpha and beta versions, the final version was approved and the content was then verified and published as an Android app.

BCSzone is a guided self-management psychoeducational app comprising 4 main chapters and 40 modules as follows:

- The Breast Cancer chapter includes topics such as introduction to breast cancer, diagnostic and treatment procedures, postmastectomy care, chemotherapy, radiation therapy, exercise, and proper diet.
- The Stress Management chapter addresses topics such as anxiety symptoms and stress complications, teaching the techniques of stress management and emotion management, thought stopping, diaphragmatic and conscious breathing, progressive muscle relaxation, and guided imagery.
- The Self-Esteem chapter addresses different methods of promoting self-esteem such as thoughts of self-confirmation, styles of establishing effective communication, and assertiveness.
- The Anger Management chapter addresses anger management techniques and problem solving.

The main page of the app includes keys to the image gallery, objectives, about us, and references, which could be selected by the user to switch between the pages and use the desired content of the app.

The objectives part of the app includes the topics and their duration in a way that users could get to know about the courses, their order, and the time required to learn the contents of each part. The keys “Start lesson” and “Exit” are included at the bottom of the main page of the app.

When the Start button is clicked, the lessons begin to be run in a multimedia format and a speaker’s voice is played to help users select the features of the app including stop, forward, backward, and the next or previous slide. Each topic contains pictures, texts, tables, audio files, and animations tailored to the content, and the users must correctly respond to a multiple-choice test at the end of each educational topic before being permitted to continue studying and observing future slides. Users are provided with feedback if they incorrectly answer the questions. Efforts were therefore made to turn the teaching/learning process into a dynamic and effective process.

Procedure

BCSzone was installed on the mobile phones of the participants in the intervention group, and they were individually taught to use the app. They were also provided with written instructions and compact discs containing the app. During the study, the intervention group was provided with a phone number to contact if they encountered any technical problems associated with BCSzone.

After the app was installed on participant smartphones, educational materials could be viewed in a multimedia format with a test at the end of each educational topic. Participants who could learn the contents and consequently pass the tests in each chapter were directed to a WhatsApp online chat group to interact with one another. As the most popular communication and social media platform in Iran, WhatsApp constitutes a smartphone-based instant messaging app that enables users to send voice, text, video, or location using group communication features. A psychiatric nurse followed up with the patients and provided them with support, advice, and guidance for about 60 minutes per weekend, if needed, through the same WhatsApp online group chat. As the only reminder, the patients were reminded by the psychiatric nurse to use the app during the weekly sessions.

As found in Paxling et al [47], the psychiatric nurse feedback was based on guidelines as follows: validating the work of the participants and reporting difficulties, problem solving and clarification on how to perform the treatment methods, promoting the tasks, reinforcing the progress, and encouraging work continuation. The intervention was used by the intervention group during 4 weeks. One week after the intervention, study participants were asked to recomplete the STAI and RSES. The

control group on the waiting list received routine health care only and completed the questionnaires.

Statistical Analyses

All completed questionnaires were coded, and the data collected were analyzed in SPSS (version 22, IBM Corp). The demographic variables were expressed using descriptive statistics, including mean, standard deviation, and frequency. Depending on the type of data, 1-way analysis of variance, Fisher exact test, independent *t* test, paired *t* test, or chi-square test were used to perform inferential analysis. Differences between the 2 groups were examined by analyzing the categorical data using the Pearson chi-square test or Fisher exact test, the continuous normally distributed data using the *t* test, and the nonnormally distributed data using the Wilcoxon Mann-Whitney test. The Kolmogorov-Smirnov test was also used to investigate the data normality. $P<.05$ was set as the level of statistical significance.

Given the normal distribution of the variables, the paired *t* test was used to compare changes in the mean scores of total anxiety, state anxiety, trait anxiety, and self-esteem in the women with breast cancer in each group between before and 1 week after the psychoeducational intervention. The 2 groups were also compared using the independent *t* test.

The effect size of the intervention was estimated at small for Cohen $d=0.20$, medium for Cohen $d\sim0.50$ and large for Cohen $d\geq0.80$ [48]. In addition, no missing values were reported in the main outcomes.

Results

Recruitment and Follow-up

This trial was conducted between October 2016 and February 2017. Recruitment and baseline assessments were first performed for 3 months, the intervention for 4 weeks, and the postintervention assessments for 1 week after the intervention. In winter 2016, recruitment was performed face-to-face during the clinic's health appointment in two breast clinics affiliated with Shiraz University of Medical Sciences, Shiraz, Iran. A total of 82 women with breast cancer were randomly assigned to the intervention or wait-list control groups of 41. The flowchart in Figure 1 shows the participants included in each group, the excluded ones, and those who dropped out and their reason of exclusion.

Demographic Characteristics

The demographic characteristics for all study participants are displayed in Table 1. Results showed nonsignificant differences between the 2 groups in terms of level of education (Fisher exact test, $P=.85$), marital status (Fisher exact test, $P=.69$), history of mastectomy (chi-square test, $P=.65$), age ($P=.66$), and employment status (Fisher exact test, $P=.82$) before the intervention. The mean age of the participants was 46.45 (SD 9.29) years, that of the control group 46 (SD 8.80) years and that of the intervention group 46.9 (SD 9.83) years. The majority of the participants were married (66/82, 81%) and housewives (65/82, 79%).

Table 1. Frequency distribution of the demographic characteristics of the study groups.

Characteristics	Intervention group (n=41)	Control group (n=41)	Total (n=82)	*P* value
Age (years), mean (SD)	46.9 (9.83)	46 (8.80)	46.45 (9.29)	.66[a]
Education level, n (%)				.85[b]
Middle school	19 (46)	22 (54)	41 (50)	
Diploma	18 (44)	14 (34)	32 (39)	
Associate degree	1 (2)	1 (2)	2 (2)	
Bachelor's degree	3 (7)	4 (10)	7 (9)	
Marital status, n (%)				.69[b]
Single	2 (5)	3 (7)	5 (6)	
Married	35 (85)	31 (76)	66 (81)	
Widowed	3 (7)	5 (12)	8 (10)	
Divorced	1 (2)	2 (5)	3 (4)	
Job status, n (%)				.82[b]
Staff/worker	5 (12)	7 (17)	12 (15)	
Retired	2 (5)	2 (5)	4 (5)	
Student	0 (0)	1 (2)	1 (1)	
Housewife	34 (83)	31 (76)	65 (79)	
Mastectomy, n (%)				.65[c]
Yes	23 (56)	26 (63)	49 (60)	
No	18 (44)	15 (37)	33 (40)	

[a]Independent sample *t* test.

[b]Fisher exact test.

[c]Chi-square test.

Anxiety

The intervention and control groups were not statistically different at baseline in terms of the mean scores of total anxiety (*P*=.41), state anxiety (*P*=.35), and trait anxiety (*P*=.65). Statistically significant differences were, however, observed between the 2 groups 1 week after completing the intervention (*P*<.001). Comparing the postintervention mean score of anxiety with its preintervention mean score in the intervention group using the paired *t* test showed significant reductions in the total score of anxiety (95% CI –17.44 to –8.90, *P*<.001) and the scores of state anxiety (95% CI –9.20 to –4.21, *P*<.001) and trait anxiety (95% CI –8.50 to –4.12, *P*<.001), confirming the effectiveness of the psychoeducational intervention in alleviating anxiety in the intervention group (Table 2).

Table 2. Comparison of changes in the mean scores of anxiety and self-esteem in the women in the intervention and control groups before and after the intervention.

Variable	Pretest mean (SD)	Posttest mean (SD)	Mean differences (95% CI)	*P* value	*t* score	Effect size
Anxiety (total)						
G1[a]	103.68 (14.73)	90.66 (13.84)	–13.02 (–17.44 to –8.90)	<.001	6.40	1.02
G2[b]	106.31 (16.36)	106.92 (15.94)	0.61 (–0.69 to 1.92)	.34	0.95	0.15
State anxiety						
G1	51.39 (10.16)	44.68 (8.26)	–6.71 (–9.20 to –4.21)	<.001	5.45	0.88
G2	53.54 (8.92)	53.62 (8.65)	0.07 (–1.00 to 1.15)	.88	0.14	0.02
Trait anxiety						
G1	52.29 (7.29)	45.97 (7.97)	–6.31 (–8.50 to –4.12)	<.001	5.83	0.94
G2	52.77 (9.23)	53.31 (8.89)	0.53 (–0.35 to 1.43)	.22	1.22	0.18
Self-esteem						
G1	15.92 (4.35)	17.97 (4.69)	2.05 (1.28 to 2.82)	<.001	5.39	0.87
G2	16.64 (3.68)	16.87 (3.48)	0.23 (–0.24 to 0.71)	.33	0.97	0.15

[a]G1: intervention group.

[b]G2: control group.

Self-Esteem

No statistically significant differences were observed in the baseline mean scores of self-esteem in the 2 groups (*P*=.82). Comparing the postintervention mean scores of self-esteem showed statistically nonsignificant differences between the control and intervention groups (16.87 vs 17.97, *P*=.24).

Table 2 shows significant differences in the mean score of self-esteem between the pretest and posttest in the intervention group (95% CI –2.82 to –1.28, *P*<.001) with a large effect size (*d*=0.87) compared with those in the control group (95% CI –0.71 to 0.28, *P*=.33, *d*=0.15). In other words, the increase observed in the self-esteem scores after the intervention was more significant in the intervention group (95% CI –2.82 to –1.28, *P*<.001), suggesting the effectiveness of the psychoeducational intervention in increasing self–esteem in the intervention group.

In the intervention group, the within-group effect size (*d*) of all the measures was large between the pretest and posttest.

Satisfaction With the Mobile-Based Intervention

Table 3 presents the results of the survey conducted at the end of the study on the psychoeducational intervention performed in the intervention group using the mobile app, showing that 92% (35/38) of participants were very satisfied with the mobile educational app. They identified mobile app-based educational materials as cost-effective and had a great tendency to receive psychoeducational interventions through the mobile app rather than through in-person meetings. According to 92% (35/38) of participants, mobile educational apps could provide them with the required educational materials independently of time and place.

Table 3. Results of the survey of satisfaction with the mobile-based intervention.

Statement	Very frequently, n (%)	Frequently, n (%)	Occasionally, n (%)	Rarely, n (%)	Very rarely, n (%)
1. The mobile app meets my educational needs.	27 (70)	9 (24)	2 (5)	0 (0)	0 (0)
2. I am satisfied with the mobile app and how it was installed on my phone.	21 (55)	8 (21)	9 (24)	0 (0)	0 (0)
3. I tend to receive mobile-based psycho educational interventions rather than face-to-face meetings.	28 (74)	10 (26)	0 (0)	0 (0)	0 (0)
4. It's easy for me to use educational mobile apps.	18 (47)	15 (40)	4 (11)	1 (3)	0 (0)
5. Multimedia contents used in the mobile app are appropriate.	33 (87)	4 (11)	1 (3)	0 (0)	0 (0)
6. I prefer to receive educational materials through mobile apps.	27 (71)	9 (24)	1 (3)	1 (3)	0 (0)
7. Psychoeducational intervention through mobile technologies is cost effective for me.	38 (100)	0 (0)	0 (0)	0 (0)	0 (0)
8. It's easy for me to talk about the educational needs and personal issues that come with the disease through the online chat room.	19 (50)	8 (21)	4 (11)	5 (13)	2 (5)
9. Through the educational mobile app, I can access educational materials at anytime and anywhere.	35 (92)	3 (8)	0 (0)	0 (0)	0 (0)
10. I like to use mobile phones to share educational information about the disease and how to control it.	20 (53)	9 (24)	9 (24)	0 (0)	0 (0)
11. By using this mobile app, I feel I can contribute more in making decisions about self-care.	17 (45)	16 (42)	5 (13)	1 (3)	0 (0)
12. By using educational mobile health apps, I feel that I have more control over myself.	10 (26)	11 (29)	13 (34)	3 (8)	1 (3)
13. Getting educational materials through the mobile app about the disease and how to deal with complications is motivating me to take care of myself.	18 (47)	16 (42)	3 (8)	1 (3)	0 (0)
14. Getting information and tips about the disease and its complications through the mobile app helps save time.	37 (90)	4 (11)	0 (0)	0 (0)	0 (0)
15. I can get information and guidance on the illness and how to deal with it through the mobile app when health care and education services are not available.	36 (95)	2 (5)	0 (0)	0 (0)	0 (0)
16. I am satisfied with the educational mobile app in general.	35 (92)	3 (8)	0 (0)	0 (0)	0 (0)

Discussion

Principal Findings

To the best of our knowledge, this clinical trial pioneered the investigation of the effect of mobile-based psychoeducation on women with breast cancer. Our study hypothesis stated mobile-based psychoeducation can increase self-esteem and decrease anxiety in patients with breast cancer. The intervention group received the psychoeducational intervention using BCSzone and participated in nurse-assisted online mobile support sessions. The hypothesis suggesting decreases in anxiety scores was confirmed in the intervention group. After the first week of the intervention, the scores reported for state anxiety and trait anxiety in the intervention were statistically and significantly higher than in the control group.

These findings suggest the effectiveness of psychoeducational interventions through a mobile app (BCSzone) in reducing state anxiety and trait anxiety in women with breast cancer, which

is consistent with literature on the effectiveness of psychoeducational interventions that focused on patients with breast cancer delivered via group, face-to-face, and/or video formats [14-16,49].

The positive results obtained cannot be explained by a single factor, as psychoeducational interventions encompass a wide range of trainings and mental behavioral skills and psychological techniques. The obtained results can be explained by the factors used in this study such as cognitive behavioral methods, including relaxation and adaptation skills, thought stopping, conscious breathing, guided imagery, and stress and emotional management, whose effectiveness in emotional disorders such as anxiety has been proven [50,51].

This study found a significant increase in self-esteem levels in the intervention group 1 week after the intervention compared with the baseline levels of self-esteem, although between-group analysis suggested nonsignificant increases in these levels in the control group. Research suggests patients with breast cancer tend to become socially isolated through limiting their own social relationships and losing their interest in group and social activities. These patients can degrade their self-esteem through social isolation, which manifests itself as failing to express their feelings, dissatisfaction with their appearance and changes in their performance, and a lack of self-confidence [3]. BCSzone addresses different self-esteem promotion methods and skills, including self-confirmation thoughts, effective communication, and assertiveness skills.

Research suggests promoting assertiveness skills and self-confirmation can assist the patients in improving their performance and social interactions [52]. Moreover, self-esteem can be enhanced by improving performance, interpersonal relationship skills, and self-satisfaction. It is also suggested that self-esteem and performance affect each other to some extent and self-esteem increases after achieving a goal. As embedded features of the mobile app, lessons on assertiveness, anger and stress management, and effective communication skills for women with breast cancer were also found to enhance their self-esteem [53].

Although between-group comparison suggested no statistically significant differences in self-esteem, the increase in self-esteem scores in the intervention group was not negligible after the intervention compared with baseline. As a complicated multidimensional disorder, low self-esteem can be treated using different interventions such as psychoeducation and cognitive behavioral therapy. Although the majority of these dimensions were discussed within the app, the limited duration of the intervention can explain the statistically nonsignificant difference between the groups.

In contrast to previous studies, this study used a smartphone app with multiple advantages to perform psychoeducational interventions. The effectiveness and efficacy of mobile and internet interventions have been recently supported in literature. These interventions were found more effective than conventional treatments and as effective as face-to-face therapies for a wide range of conditions such as depression and anxiety [21,22]. Nowadays, a significant amount of time is spent using smartphones daily, and mobile apps can be increasingly employed for improving health and managing diseases given their low cost, accessibility, and availability. All multimedia contents provided by smartphones such as video clips, text messages, and images can also assist users in improving their health [54]. Patients with cancer were found to be interested in using mobile technologies for managing their disease, and the potential of smartphones for improving patient awareness of emotional well-being and stress by delivering therapeutic interventions and reducing anxiety has been highlighted in literature [55]. Moreover, the advantages offered by online support groups compared with those offered by face-to-face groups include greater scheduling convenience and an increased access to care for those who were otherwise unable to participate in these groups due to their social anxiety, personal concerns, residence in remote areas, or general health status [56]. Recent reviews revealed that social networking apps can promote discussion, consultation, and collaboration among health care professionals and patients. This research states that online mobile-based systems enhance different psychosocial and quality of life outcomes and reduce anxiety levels in patients with breast cancer [32]. These findings are consistent with the literature, including an uncontrolled study by Attai et al [57] that reported a rise in perceived knowledge and reductions in anxiety levels in patients with breast cancer as a result of participating in a support group on Twitter.

Despite the large number of currently available telehealth and mHealth apps in oncological settings, to the best of our knowledge, the number of RCTs evaluating app effectiveness in women with breast cancer is limited. Evaluating the effect of an mobile app on women with breast cancer through an uncontrolled study, Kuijpers et al [58] reported significant improvements in the 3 dimensions of quality of life (role functioning, emotional-mental health, and physical functioning) over time. A simple app-based e-support program was also found to improve quality of life, self-efficacy, and symptom interference in women with breast cancer during their chemotherapy [59]. Moreover, a study by Zhu et al [60] showed the participants to perceive the program as effective in improving their knowledge, emotional well-being, and confidence.

Patients with breast cancer can remotely participate in interventions using platforms such as mobile devices. A pilot study by Lengacher et al [26] showed a 6-week mobile mindfulness - based program including yoga, sitting and walking meditation, and body scan to alleviate the psychophysical symptoms of stress, state anxiety, and depression and enhance quality of life in breast cancer survivors. This program was therefore found to be acceptable and feasible given its clinical effects on the psychophysical symptoms.

Numerous studies have demonstrated the impact of mobile phones and apps on the provision of support for cancer survivors and the treatment, control, and prevention of diverse disorders and diseases [18,29,61]. In contrast, Zhu et al [59] reported the insignificant effects of mobile-based support programs on depression, anxiety, and social support in women with breast cancer undergoing chemotherapy. Similarly, Foley et al [27] found an mHealth app developed to provide ample information for patients with breast cancer to negatively affect them and

exacerbate their anxiety after surgery, although they reported lower anxiety levels in members of the control group, who did not use the app. This reduction in anxiety was also found to be correlated with higher quality of life in the controls. The discrepancies in findings between this and previous research can be explained by differences in the study design, the mobile app employed, the method of teaching patients and their demographic details, and outcome measures.

Educational subjects and practical and psychological tests and exercises are integrated into BCSzone. Each module ends by providing users with feedback on their incorrect responses to the test questions, if any. MacDonald [62] found the possibility of interacting and receiving feedback from the therapist, different features of the apps and homework, and symptom distress tracking to be the factors that help enhance patient adherence to mHealth apps. The cyberspace relationship also helps patients with free expression of their personal problems and sharing of their experiences and feelings. Smartphone and mobile apps provide a great opportunity to discover new modalities of monitoring, treatment, and research into psychiatric and mental health conditions [63]. The promising platform provided by mobile apps helps women with breast cancer acquire knowledge and interact with peers and health care experts whenever and wherever they need to [20].

This research found the intervention based on BCSzone as a multicomponent psychoeducation mobile app and a nurse-administered online group to enhance self-esteem and decrease anxiety in patients with breast cancer.

These findings can be improved by employing diverse interventions, including psychiatric nurse feedback and participation in online group discussion. Research suggests advanced social media and communication technology can be used to increase the involvement of patients with cancer in diverse types of discussion and reflection on their own health [64]. In addition, mobile and web-based online support groups enable women with breast cancer to express psychophysical and sexual problems arising from their disease and easily share their feelings and experiences with therapists and other patients; otherwise, they would be hesitant to express their problems in face-to-face support groups [65]. In line with our research, a study by Winzelberg at al [56] found an internet support group to help reduce perceived stress, cancer-associated traumas, and depression in females with primary breast cancer.

Seeking to share information, feelings, and coping strategies was reported as the main inclination of patients with breast cancer. Online communities constitute reliable platforms for providing cancer survivors with opportunities to meet other survivors, exchange ideas, and receive support [66].

According to McGraw and Hall [67], serious security and privacy issues in mobile apps and telehealth systems serve as obstacles to the clinical and industrial applications of these systems; nevertheless, many populations, especially chronically ill patients, have reported that the beneficial effects of using telehealth systems and mHealth apps outweigh the risks [68]. This study included only the psychiatric nurse and study participants in a private WhatsApp group, and the messages were end-to-end encrypted to ensure privacy in sharing identifiable health data.

At the end of this study, the intervention group was asked to complete the author-designed survey assessing their opinions of and satisfaction with the mobile app. The results showed satisfaction with the mobile app in 92% of the cases. They all believed in the cost-effectiveness of receiving educational materials through mobile apps and were eager to electronically receive psychoeducational interventions through mobile phones rather than through personal meetings. They also identified the use of mobile educational apps as a new experience. Using BCSzone was therefore found to help with the stepwise care of patients with cancer.

In approximately half the cases, patients with cancer were recently found to be willing to send and receive data through an app supporting their oncological treatment and follow-ups [23]. In line with this research, a study on mobile-based education and interventions reported that 69.2% of the participants found mobile phones to be effective, simple, and immediate tools in their learning, 72.2% found mobile-aided learning to be a new experience, and 73.4% believed that the mobile-aided learning approach is learner-oriented and flexible in time, pace, and place. Educational outcomes can be improved through developing the essential functionality of certain teaching and learning techniques based on the unique features of mobile phones. Furthermore, mobile systems can be employed to raise patient levels of access to psychological support and remotely perform psychological interventions when patients are vulnerable [69,70].

Practical Implications

These findings can be used in evidence-based practices to determine the efficacy and cost-effectiveness of interventions in improving patient outcomes. Nurses and other health professionals can employ mobile technology to provide a larger number of patients with care, counseling services, and trainings. Mobile educational apps are important for encouraging women with breast cancer to participate in self-care programs. Given the positive and useful results obtained from administering the psychoeducational intervention using mobile apps and mobile-based online group discussions, it is recommended that this intervention be used as a simple, cost-effective, and useful method in patients with breast cancer.

Study Limitations

This study limitations included its short duration (4 weeks) given that psychoeducation should be regularly performed over a long period. Moreover, several health care workers and more follow-ups are required for evaluating the long-term effects of the mobile app. Further studies with longer follow-ups and larger samples are also recommended to be conducted in this context. Furthermore, this study failed to measure adherence to the mobile app and the length of use by the users.

Conclusion

These findings are consistent with the results of previous studies on patient education and provision of counseling and support through mobile apps. BCSzone and the nurse-administered

online group discussion were found to reduce anxiety levels and raise self-esteem scores in women with breast cancer. This research highlighted the key role of using mobile apps in psychoeducational interventions in decreasing anxiety and improving self-esteem in women with breast cancer. Further studies are recommended that be conducted in this context using longer follow-ups.

Acknowledgments

The authors would like to express their gratitude to the authorities of Shiraz University of Medical Sciences for approving and financially supporting this study as part of a dissertation thesis (grant 94-01-08-10534).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey of Satisfaction with the Mobile-Based Intervention.
[PDF File (Adobe PDF File), 157 KB - mhealth_v9i5e19262_app1.pdf]

Multimedia Appendix 2
Screenshots of the mobile app (BCSzone).
[PNG File , 578 KB - mhealth_v9i5e19262_app2.png]

Multimedia Appendix 3
CONSORT-EHEALTH checklist V1.6.
[PDF File (Adobe PDF File), 2594 KB - mhealth_v9i5e19262_app3.pdf]

References

1. Jazayeri SB, Saadat S, Ramezani R, Kaviani A. Incidence of primary breast cancer in Iran: ten-year national cancer registry data report. Cancer Epidemiol 2015 Aug;39(4):519-527. [doi: 10.1016/j.canep.2015.04.016] [Medline: 26070507]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7-30 [FREE Full text] [doi: 10.3322/caac.21332] [Medline: 26742998]
3. Al-Azri M, Al-Awisi H, Al-Rasbi S, El-Shafie K, Al-Hinai M, Al-Habsi H, et al. Psychosocial impact of breast cancer diagnosis among Omani women. Oman Med J 2014 Nov;29(6):437-444 [FREE Full text] [doi: 10.5001/omj.2014.115] [Medline: 25584162]
4. Saniah A, Zainal N. Anxiety, depression and coping strategies in breast cancer patients on chemotherapy. Malaysian J Psychiatr 2010;19(2):1-6 [FREE Full text]
5. Alacacioglu A, Yavuzsen T, Dirioz M, Yilmaz U. Quality of life, anxiety and depression in Turkish breast cancer patients and in their husbands. Med Oncol 2009 Dec;26(4):415-419. [doi: 10.1007/s12032-008-9138-z] [Medline: 19031014]
6. Gold M, Dunn LB, Phoenix B, Paul SM, Hamolsky D, Levine JD, et al. Co-occurrence of anxiety and depressive symptoms following breast cancer surgery and its impact on quality of life. Eur J Oncol Nurs 2016 Feb;20:97-105 [FREE Full text] [doi: 10.1016/j.ejon.2015.06.003] [Medline: 26187660]
7. Lim CC, Devi MK, Ang E. Anxiety in women with breast cancer undergoing treatment: a systematic review. Int J Evid Based Healthc 2011 Sep;9(3):215-235. [doi: 10.1111/j.1744-1609.2011.00221.x] [Medline: 21884450]
8. Enache RG. The relationship between anxiety, depression and self-esteem in women with breast cancer after surgery. Procedia Soc Behav Sci 2012;33:124-127. [doi: 10.1016/j.sbspro.2012.01.096]
9. Yektatalab S, Taleii A, Moosavinasab M, Soleimani S. Sexual dysfunction in breast cancer survivors. J Client Centered Nurs Care 2015;1(1):9-14 [FREE Full text] [doi: 10.32598/jccnc.1.1.9]
10. Prates ACL, Freitas-Junior R, Prates MFO, Veloso MDF, Barros NDM. Influence of body image in women undergoing treatment for breast cancer. Rev Bras Ginecol Obstet 2017 Apr;39(4):175-183 [FREE Full text] [doi: 10.1055/s-0037-1601453] [Medline: 28359110]
11. Yektatalab S, Ghanbari E. The relationship between anxiety and self-esteem in women suffering from breast cancer. J Midlife Health 2020;11(3):126-132 [FREE Full text] [doi: 10.4103/jmh.JMH_140_18] [Medline: 33384534]
12. Sowislo JF, Orth U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. Psychol Bull 2013 Jan;139(1):213-240. [doi: 10.1037/a0028931] [Medline: 22730921]
13. Cruickshank S, Kennedy C, Lockhart K, Dosser I, Dallas L. Specialist breast care nurses for supportive care of women with breast cancer. Cochrane Database Syst Rev 2008 Jan 23(1):CD005634. [doi: 10.1002/14651858.CD005634.pub2] [Medline: 18254086]

14. Cozaru GC, Papari AC, Sandu ML. The effects of psycho-education and counselling for women suffering from breast cancer in support groups. Procedia Soc Behav Sci 2014 Apr;128:10-15. [doi: 10.1016/j.sbspro.2014.03.110]
15. Arafa ME, Hassan M. Psychoeducational program for breast cancer survivors, effect on cancer-related fatigue and quality of life. Egypt J Psychiatr 2013;34(1):25-33 [FREE Full text] [doi: 10.7123/01.EJP.0000415299.41177.69]
16. Dastan NB, Buzlu S. Psychoeducation intervention to improve adjustment to cancer among Turkish stage I-II breast cancer patients: a randomized controlled trial. Asian Pac J Cancer Prev 2012;13(10):5313-5318 [FREE Full text] [doi: 10.7314/apjcp.2012.13.10.5313] [Medline: 23244155]
17. Ryu S. Book Review: mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231-233. [doi: 10.4258/hir.2012.18.3.231]
18. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12(1):67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
19. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Optimizing cancer care through mobile health. Support Care Cancer 2015 Jul;23(7):2183-2188. [doi: 10.1007/s00520-015-2627-7] [Medline: 25649121]
20. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
21. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Human Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
22. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
23. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile health in oncology: a patient survey about app-assisted cancer care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
24. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: a randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
25. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and phone-based weight loss intervention: pilot RCT in women at elevated breast cancer risk. Am J Prev Med 2016 Nov;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]
26. Lengacher CA, Reich RR, Ramesar S, Alinat CB, Moscoso M, Cousin L, et al. Feasibility of the mobile mindfulness-based stress reduction for breast cancer (mMBSR(BC)) program for symptom improvement among breast cancer survivors. Psychooncology 2018 Feb;27(2):524-531 [FREE Full text] [doi: 10.1002/pon.4491] [Medline: 28665541]
27. Foley NM, O'Connell EP, Lehane EA, Livingstone V, Maher B, Kaimkhani S, et al. PATI: Patient accessed tailored information: a pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]
28. Ringwald J, Marwedel L, Junne F, Ziser K, Schäffeler N, Gerstner L, et al. Demands and needs for psycho-oncological ehealth interventions in women with cancer: cross-sectional study. JMIR Cancer 2017 Nov 24;3(2):e19 [FREE Full text] [doi: 10.2196/cancer.7973] [Medline: 29175813]
29. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
30. Chow PI, Showalter SL, Gerber M, Kennedy EM, Brenin D, Mohr DC, et al. Use of mental health apps by patients with breast cancer in the united states: pilot pre-post study. JMIR Cancer 2020 Apr 15;6(1):e16476 [FREE Full text] [doi: 10.2196/16476] [Medline: 32293570]
31. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
32. Pope Z, Lee JE, Zeng N, Lee HY, Gao Z. Feasibility of smartphone application and social media intervention on breast cancer survivors' health outcomes. Transl Behav Med 2019 Jan 01;9(1):11-22. [doi: 10.1093/tbm/iby002] [Medline: 29471477]
33. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S467-S472 [FREE Full text] [doi: 10.1002/acr.20561] [Medline: 22588767]
34. Rosenberg M, Schooler C, Schoenbach C. Self-esteem and adolescent problems: modeling reciprocal effects. Am Sociol Rev 1989 Dec;54(6):1004-1018. [doi: 10.2307/2095720]
35. Aghabarari M, Ahmadi F, Aghaalinejad H, Mohammadi E, Hajizade E. Effect of designed exercise program on stress, anxiety and depression in women with breast cancer receiving chemotherapy. J Shahrekord Univ Med Sci 2008;9(4):26-35 [FREE Full text]

36. The effect of psycho-educational interventions through mobile technologies on self-esteem and anxiety in women with breast cancer. Iranian Registry of Clinical Trials. URL: https://en.irct.ir/trial/19882 [accessed 2021-05-05]
37. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. J Clin Psychiatry 1983;53:267-271 [FREE Full text] [doi: 10.1002/9780470479216.corpsy0943]
38. Spielberger CD. State-Trait Anxiety Inventory. In: The Corsini Encyclopedia of Psychology. Hoboken: John Wiley & Sons; 2010. [doi: 10.1002/9780470479216.corpsy0943]
39. Mahram B. [Manual of Spielberger State-Trait Anxiety Inventory] [Thesis]. Mashhad: Ferdowsi University of Mashhad; 1993:52-54.
40. Abdoli N, Farnia V, Salemi S, Davarinejad O, Ahmadi Jouybari T, Khanegi M, et al. Reliability and validity of Persian version of State-Trait Anxiety Inventory among high school students. East Asian Arch Psychiatr 2020 Jun;30(2):44-47 [FREE Full text] [doi: 10.12809/eaap1870] [Medline: 32611826]
41. Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. J Pers Soc Psychol 2005 Oct;89(4):623-642. [doi: 10.1037/0022-3514.89.4.623] [Medline: 16287423]
42. Shapurian R, Hojat M, Nayerahmadi H. Psychometric characteristics and dimensionality of a Persian version of Rosenberg Self-esteem Scale. Percept Mot Skills 1987 Aug;65(1):27-34. [doi: 10.2466/pms.1987.65.1.27] [Medline: 3684462]
43. Alizadeh T, Farahani M, Shahraray M, Alizadegan S. The relationship between self esteem and locus of control with infertility related stress of no related infertile men and women. J Reprod Infertil 2005;6(2):194-204 [FREE Full text]
44. Jafari F, Khatony A, Mehrdad M. Self-Esteem Among the Elderly Visiting the Healthcare Centers in Kermanshah-Iran (2012). Glob J Health Sci 2015 Apr 15;7(5):352-358 [FREE Full text] [doi: 10.5539/gjhs.v7n5p352] [Medline: 26156932]
45. Fani M, Mehravar S, Mehrabi M. Level of learning and satisfaction through traditional methods and the use of multimedia: a comparative study. Interdiscipl J Virtual Learning Med Sci 2014;5(2):72-78 [FREE Full text]
46. Mehrabi M. Setting up the structure and process for e-content development. Interdiscipl J Virtual Learning Med Sci 2019;10(4):78-80. [doi: 10.30476/ijvlms.2019.84733.1016]
47. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]
48. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Edition. Hillsdale: Erbaum Press; 1988. [doi: 10.4324/9780203771587]
49. Izadi-Ajirlo A, Bahmani B, Ghanbari-Motlagh A. Effectiveness of cognitive behavioral group intervention on body image improving and increasing self-esteem in women with breast cancer after mastectomy. J Rehab 2013;13(4):72-83 [FREE Full text]
50. Torabizadeh C, Bostani S, Yektatalab S. Comparison between the effects of muscle relaxation and support groups on the anxiety of nursing students: A randomized controlled trial. Complement Ther Clin Pract 2016 Nov;25:106-113. [doi: 10.1016/j.ctcp.2016.09.001] [Medline: 27863599]
51. Yektatalab S, Khodadadi S, Moattari M, Hosseiny FN, Zare N. A web-based anger management program for parent-female adolescents' conflicts: a cluster randomized controlled trial. Int J Adolesc Med Health 2017 Jul 18;31(3):30. [doi: 10.1515/ijamh-2017-0012] [Medline: 28719366]
52. Fitriasari A, Yusuf A, Probowati R. Assertive community treatment of psychosocial distress in patients with breast cancer. Int J Sci Res 2017 Jul 05;6(7):253-256. [doi: 10.21275/art20174780]
53. Hinnen C, Hagedoorn M, Ranchor AV, Sanderman R. Relationship satisfaction in women: a longitudinal case-control study about the role of breast cancer, personal assertiveness, and partners' relationship-focused coping. Br J Health Psychol 2008 Nov;13(Pt 4):737-754. [doi: 10.1348/135910707X252431] [Medline: 17999780]
54. Heo J, Chun M, Lee KY, Oh Y, Noh OK, Park RW. Effects of a smartphone application on breast self-examination: a feasibility study. Healthc Inform Res 2013 Dec;19(4):250-260 [FREE Full text] [doi: 10.4258/hir.2013.19.4.250] [Medline: 24523989]
55. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of mobile devices to help cancer patients meet their information needs in non-inpatient settings: systematic review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026 [FREE Full text] [doi: 10.2196/10026] [Medline: 30552082]
56. Winzelberg AJ, Classen C, Alpers GW, Roberts H, Koopman C, Adams RE, et al. Evaluation of an internet support group for women with primary breast cancer. Cancer 2003 Mar 01;97(5):1164-1173 [FREE Full text] [doi: 10.1002/cncr.11174] [Medline: 12599221]
57. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
58. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]

59. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
60. Zhu J, Ebert L, Guo D, Yang S, Han Q, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 1): qualitative study of women's perceptions. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e85 [FREE Full text] [doi: 10.2196/mhealth.9311] [Medline: 29643056]
61. Cruz FOAM, Vilela RA, Ferreira EB, Melo NS, Reis PEDD. Evidence on the use of mobile apps during the treatment of breast cancer: systematic review. JMIR Mhealth Uhealth 2019 Aug 27;7(8):e13245 [FREE Full text] [doi: 10.2196/13245] [Medline: 31456578]
62. MacDonald J, Amoyal N, Nisotel L, Perez GK, Traeger L, Temel JS, et al. Mobile application of cognitive behavioral therapy for anxiety and advanced cancer. J Clin Oncol 2015 Oct 10;33(29_suppl):220-220. [doi: 10.1200/jco.2015.33.29_suppl.220] [Medline: 28148086]
63. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
64. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
65. Vilhauer RP. Perceived benefits of online support groups for women with metastatic breast cancer. Women Health 2009;49(5):381-404. [doi: 10.1080/03630240903238719] [Medline: 19851944]
66. van Eenbergen MC, van de Poll-Franse LV, Heine P, Mols F. The impact of participation in online cancer communities on patient reported outcomes: systematic review. JMIR Cancer 2017 Sep 28;3(2):e15 [FREE Full text] [doi: 10.2196/cancer.7312] [Medline: 28958985]
67. Hall JL, McGraw D. For telehealth to succeed, privacy and security risks must be identified and addressed. Health Aff (Millwood) 2014 Feb;33(2):216-221. [doi: 10.1377/hlthaff.2013.0997] [Medline: 24493763]
68. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
69. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
70. Absavaran M, Niknami S, Zareban I. Effect of training through lecture and mobile phone on Breast Self-Examination among nurses of Zabol Hospitals. Payesh J 2015;14(3):363-373 [FREE Full text]

Abbreviations

BCSzone: Breast Cancer Support zone
CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
RCT: randomized controlled trial
RSES: Rosenberg Self-Esteem Scale
STAI: State-Trait Anxiety Inventory

Edited by L Buis; submitted 10.04.20; peer-reviewed by P Copanitsanou, G Perez, A Teles; comments to author 01.09.20; revised version received 02.01.21; accepted 19.03.21; published 18.05.21.

Please cite as:
Ghanbari E, Yektatalab S, Mehrabi M
Effects of Psychoeducational Interventions Using Mobile Apps and Mobile-Based Online Group Discussions on Anxiety and Self-Esteem in Women With Breast Cancer: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e19262
URL: https://mhealth.jmir.org/2021/5/e19262
doi:10.2196/19262
PMID:34003138

Original Paper

Associations Between Rural or Urban Status, Health Outcomes and Behaviors, and COVID-19 Perceptions Among Meditation App Users: Longitudinal Survey Study

Nishat Bhuiyan[1], PhD; Megan Puzia[2], MS; Chad Stecher[1], PhD; Jennifer Huberty[1], PhD

[1]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[2]Behavioral Research and Analytics, LLC, Salt Lake City, UT, United States

Corresponding Author:
Nishat Bhuiyan, PhD
College of Health Solutions
Arizona State University
500 North 3rd Street
Phoenix, AZ, 85004
United States
Phone: 1 480 406 2644
Email: nbhuiyan@asu.edu

Abstract

Background: Rural and urban differences in health outcomes and behaviors have been well-documented, with significant rural health disparities frequently highlighted. Mobile health (mHealth) apps, such as meditation apps, are a novel method for improving health and behaviors. These apps may be a critical health promotion strategy during the COVID-19 pandemic and could potentially be used to address rural health disparities. However, limited research has assessed whether meditation app health outcomes are associated with rural and urban residence, and it is unclear whether disparities in health and behaviors between rural and urban populations would persist among meditation app users.

Objective: We aimed to explore associations between rural or urban status, psychological outcomes, and physical activity among users of a mobile meditation app. We further aimed to explore associations between rural or urban status and perceived effects of COVID-19 on stress, mental health, and physical activity, and to explore changes in these outcomes in rural versus urban app users over time.

Methods: This study was a secondary analysis of a national survey conducted among subscribers to the meditation app Calm. Eligible participants completed online baseline surveys from April to June 2020, and follow-up surveys from June to September 2020, assessing demographics, psychological outcomes, physical activity, and perceived effects of COVID-19 on stress, mental health, and physical activity.

Results: Participants (N=8392) were mostly female (7041/8392, 83.9%), non-Hispanic (7855/8392, 93.6%), and White (7704/8392, 91.8%); had high socioeconomic status (income ≥US $100,000: 4389/8392, 52.3%; bachelor's degree or higher: 7251/8392, 86.4%); and resided in a metropolitan area core (rural-urban commuting area code 1: 7192/8392, 85.7%). Rural or urban status was not associated with baseline stress, depression, anxiety, pre–COVID-19 and current physical activity, or perceived effects of COVID-19 on stress, mental health, and physical activity. Repeated-measures models showed overall decreases in depression, anxiety, and perceived effects of COVID-19 on physical activity from baseline to follow-up, and no significant changes in stress or perceived effects of COVID-19 on stress and mental health over time. Models also showed no significant main effects of rural or urban status, COVID-19 statewide prevalence at baseline, or change in COVID-19 statewide prevalence.

Conclusions: We did not find associations between rural or urban status and psychological outcomes (ie, stress, depression, and anxiety), physical activity, or perceived effects of COVID-19 on stress, mental health, and physical activity. Rural or urban status does not appear to drive differences in outcomes among meditation app users, and the use of mHealth apps should continue to be explored as a health promotion strategy in both rural and urban populations. Furthermore, our results did not show negative cumulative effects of COVID-19 on psychological outcomes and physical activity among app users in our sample, the majority of whom were urban, White, female, and of high socioeconomic status. Further research is needed to investigate meditation app use as a health promotion strategy in rural and urban populations.

(JMIR Mhealth Uhealth 2021;9(5):e26037) doi:10.2196/26037

KEYWORDS

mHealth; rural health; physical activity; mental health; COVID-19

Introduction

Significant rural and urban status–related differences in health outcomes and behaviors have been well-documented in the United States [1-6]. Some studies have demonstrated poorer outcomes in urban areas, such as increased risk for certain mental disorders and cancer incidence [1,2]. However, for the majority of health outcomes and behaviors, rural residents face poorer outcomes compared to their urban counterparts, including higher chronic disease incidence and mortality rates and lower diet quality and physical activity [3-6], underscoring the need for methods to address rural health disparities.

The use of mobile health (mHealth) apps has been rapidly growing as a novel method for health promotion [7,8], as strategies that include mHealth apps may be more cost-effective and scalable as well as have wider reach than in-person–delivered programs [9]. mHealth tools may be even more critical during the current COVID-19 pandemic, given the limited availability and potential risks associated with in-person–delivered programs [10]. Studies have demonstrated that mHealth apps can effectively improve mental and physical health and behaviors [7,8]. Rural areas in the United States are often characterized by limitations in resources, such as medical facilities and specialty health care services; thus, the use of mHealth apps to remotely deliver health interventions to improve outcomes in rural populations may be of paramount importance to rural communities and help to address ongoing rural health disparities [11,12].

The use of meditation apps, in particular, could also be used to potentially address disparities in mental health outcomes and health behaviors faced by rural residents. Meditation is a well-known strategy for improving mental health outcomes, such as stress, depression, and anxiety [13,14]. Furthermore, growing literature further suggests that meditation is a promising tool for counteracting sedentariness to address physical inactivity [15]. Given the poorer mental health outcomes [16,17], lower access to mental health treatment [17], and lower prevalence of positive health outcomes and behaviors among rural residents [18], research exploring rural and urban status–related differences in health and behavior outcomes among meditation app users is warranted.

Limited research has assessed whether mHealth app outcomes are associated with rural and urban residence, and it is currently unclear whether disparities in health and behaviors between rural and urban populations would persist among mobile meditation app users. Thus, the purpose of this study is to (1) explore associations between rural or urban status, psychological outcomes, and physical activity; (2) explore associations between rural or urban status and perceived effects of COVID-19 on stress, mental health, and physical activity; and (3) assess changes in psychological outcomes, physical activity, and perceived effects of COVID-19 over time among rural versus urban users of a mobile meditation app.

Methods

Overview

This study was a secondary analysis of a national survey conducted in a nonrandom convenience sample of paying subscribers to the mobile meditation app Calm. All study materials and procedures were reviewed and approved by the Institutional Review Board at Arizona State University (protocol ID: STUDY00014534). Potentially eligible subscribers met the following inclusion criteria: (1) were 18 years of age or older, (2) had opened an email from Calm and used Calm at least once in the last 90 days, (3) were able to read and understand English, and (4) were United States residents. Potentially eligible subscribers were identified by the Calm informatics team and were sent a study recruitment email by the research team. The recruitment email included a brief study description and the link to an online Qualtrics eligibility survey to verify that they met the following inclusion criteria: (1) were at least 18 years of age, (2) were able to read and understand English, and (3) resided in the United States or a United States territory. This survey was free and voluntary and took approximately 3 minutes to complete.

Eligible participants were emailed a link to complete an electronic informed consent form, which stated the study purpose; the identity of the investigator; the length of time to complete the survey; which data were stored, where, and for how long; potential risks and benefits; and compensation details. Eligible participants were emailed links to online Qualtrics baseline and follow-up surveys. Surveys took approximately 15 minutes to complete and were free and voluntary; in addition, participants were able to skip survey questions.

The time frame for this study was approximately 2 months, from baseline to follow-up. Baseline surveys were distributed from April 22 to June 3, 2020, and follow-up surveys were distributed from June 26 to September 11, 2020. To maintain participant confidentiality, study data were imported into a secure and backed-up MySQL (Structured Query Language) database. Each database user had their own username and password, with permissions set appropriately for that user. Query logging was enabled to provide an electronic audit trail that recorded all user interactions with the database. The database did not store personally identifiable information; instead, records were linked to individual participants via a unique participant ID. Regarding compensation, all participants were entered into a random drawing for 1 of 20 gift cards valued at US $50.00 for completing the baseline survey (ie, 20 participants received US $50.00 at baseline) and a random drawing for 1 of 20 gift cards valued at US $50.00 for completing the follow-up survey (ie, 20 participants received US $50.00 at follow-up).

Measures

The online Qualtrics baseline and follow-up surveys included both investigator-developed and validated questionnaires. Participants self-reported demographic characteristics; stress, via the Perceived Stress Scale [19]; depression and anxiety, via the Hospital Anxiety and Depression Scale (HADS) [20]; pre–COVID-19 and current physical activity (days/week); and the extent to which COVID-19 affected their stress, mental health, and physical activity. The Perceived Stress Scale includes 10 items that measure the degree of self-appraised stress in one's life within the past month [19]. Response items are scored on a 5-point Likert scale ranging from 0 (never) to 4 (very often). Items are summed to produce a total score ranging from 0 to 40, with higher scores indicating higher levels of perceived stress. The Perceived Stress Scale is a reliable and valid measure that has demonstrated good internal consistency (Cronbach α=.74-.91) [19]. The HADS is a 14-item scale measuring levels of anxiety and depression. Seven items comprise the anxiety subscale (HADS-A) and seven items comprise the depression subscale (HADS-D) [20]. Response items are scored on a 4-point Likert scale ranging from 0 to 3. Items are summed to produce a total score ranging from 0 to 21 for each subscale. Scores between 0 and 7 are considered normal, scores from 8 to 10 are considered borderline abnormal, and scores from 11 to 21 are considered abnormal. The HADS is a valid and reliable tool, with internal consistency reported to reach α levels of .93 and .90 for the HADS-A and HADS-D subscales, respectively [20]. Participants were asked to self-report how many days per week (on a scale of 0 to 7) of physical activity they participated in prior to COVID-19 as well as their current participation. Participants were asked, via investigator-developed items, "To what extent do you feel the COVID-19 pandemic has affected your stress?" "To what extent do you feel the COVID-19 pandemic has affected your mental health?" and "To what extent do you feel the COVID-19 pandemic has affected your physical activity?" Response items were scored on a 5-point Likert scale, ranging from 0 (not at all) to 4 (very much), with higher summed scores representing a greater impact of COVID-19 on outcomes.

Participant ZIP Codes were categorized into levels of urbanization using the US Department of Agriculture's rural-urban commuting area (RUCA) codes, which range from 1, indicating a metropolitan core area, to 10, indicating a rural area [21]. RUCA code definitions can be found in Table 1. To define rural or urban status, a binary variable was created by dummy-coding RUCA 1 as urban and RUCA 2 to 10 as rural. As an alternative approach, we also treated the RUCA code as a continuous variable in replicated study analyses, the results of which can be found in Multimedia Appendix 1.

COVID-19 prevalence data were derived from the US Centers for Disease Control and Prevention aggregate data set of daily numbers of confirmed and probable case and deaths over time [22]. This information was used to categorize each state's relative COVID-19 risk, where low-prevalence COVID-19 states were defined as those with fewer than 10,000 cases per 100,000 at the time of survey distribution and high-prevalence COVID-19 states were those with 10,000 cases or more per 100,000. At baseline, high-prevalence COVID-19 states included California, Colorado, Connecticut, Florida, Georgia, Illinois, Indiana, Louisiana, Massachusetts, Maryland, Michigan, New Jersey, New York, Ohio, Pennsylvania, Texas, Virginia, and Washington. At follow-up, high-prevalence COVID-19 states additionally included Alabama, Arkansas, Arizona, DC, Delaware, Iowa, Kansas, Kentucky, Minnesota, Missouri, Mississippi, North Carolina, Nebraska, New Mexico, Nevada, Oklahoma, Rhode Island, South Carolina, Tennessee, Utah, and Wisconsin. A binary variable, *COVID-19 state baseline*, was created based on high-prevalence COVID-19 states, and a second variable, *COVID-19 state change*, was created based on states that were low prevalence at baseline and changed to high prevalence at follow-up.

Table 1. Participant demographic characteristics.

Characteristic	Value (N=8392)
Age (years), mean (SD)	47.53 (13.83)
Gender (n=7303), n (%)	
Female	6129 (83.92)
Male	1147 (15.71)
Other	27 (0.37)
Ethnicity (n=6774), n (%)	
Hispanic	436 (6.44)
Non-Hispanic	6338 (93.56)
Race (n=7338), n (%)	
White	6586 (91.75)
Black or African American	231 (3.22)
Asian	216 (3.01)
Native American or Alaska Native	83 (1.16)
Native Hawaiian or Pacific Islander	27 (0.38)
Other	195 (2.72)
Income (US $) (n=6949), n (%)	
≤20,000	212 (3.05)
21,000-40,000	402 (5.79)
41,000-60,000	705 (10.15)
61,000-80,000	942 (13.56)
81,000-100,000	1055 (15.18)
>100,000	3633 (52.28)
Education level (n=7319), n (%)	
11^{th} grade or less	8 (0.11)
High school or General Educational Development	161 (2.20)
Some college	826 (11.29)
2-year college or technical degree	2670 (36.48)
Bachelor’s degree	424 (5.79)
Graduate degree	3230 (44.13)
Employment (n=7297), n (%)	
Employed	5084 (69.67)
Retired	1012 (13.87)
Unemployed	477 (6.54)
Homemaker	306 (4.19)
Unable to work	252 (3.45)
Student	166 (2.27)
Rural-urban commuting area (RUCA) code (n=7044) , n (%)	
1: *Metropolitan area core: primary flow within an Urbanized Area (UA)*	6038 (85.73)
2: *Metropolitan area high commuting: primary flow 30% or more to a UA*	409 (5.83)
3: *Metropolitan area low commuting: primary flow 10% to 30% to a UA*	23 (0.29)
4: *Micropolitan area core: primary flow within an Urban Cluster (UC) of 10,000 to 49,999 (large UC)*	295 (4.18)
5: *Micropolitan high commuting: primary flow 30% or more to a large UC*	42 (0.63)

Characteristic	Value (N=8392)
6: *Micropolitan low commuting: primary flow 10% to 30% to a large UC*	9 (0.11)
7: *Small town core: primary flow within a UC of 2500 to 9999 (small UC)*	121 (1.69)
8: *Small town high commuting: primary flow 30% or more to a small UC*	11 (0.24)
9: *Small town low commuting: primary flow 10% to 30% to a small UC*	11 (0.22)
10: *Rural areas: primary flow to a tract outside a UA or UC*	85 (1.19)

Statistical Analysis

Descriptive statistics and frequencies were computed to describe sample characteristics. Unadjusted and adjusted regression models were used to examine the association between rural or urban status (ie, RUCA code) and (1) psychological outcomes (ie, stress, depression, and anxiety), (2) pre–COVID-19 and current physical activity, and (3) perceived effects of COVID-19 on stress, mental health, and physical activity, controlling for demographics (ie, gender, age, ethnicity, race, income, education, employment, self-reported Calm app use, and statewide COVID-19 prevalence at baseline). To assess changes in outcomes over time, repeated-measures analyses of variance (ANOVAs) were used. Models included baseline and follow-up outcomes (ie, stress, depression, anxiety, effect of COVID-19 on stress, effect of COVID-19 on mental health, and effect of COVID-19 on physical activity) as within-subjects factors, rural and urban statuses as between-subjects factors, and gender, age, ethnicity, race, income, education, employment, self-reported Calm app use, COVID-19 state baseline (ie, high statewide COVID-19 prevalence at baseline), and COVID-19 state change (ie, change from low statewide COVID-19 prevalence at baseline to high statewide COVID-19 prevalence at follow-up) as covariates. To further assess the effect of rural or urban status and changes in statewide COVID-19 prevalence on changes in outcomes over time, we tested a three-way interaction with time, COVID-19 statewide prevalence, and rural or urban status. All statistical analyses were performed using SPSS, version 26.0 (IBM Corp), with significance inferred at $P<.05$.

Results

There were total of 8392 participants. Most participants were female (7041/8392, 83.9%), non-Hispanic (7855/8392, 93.6%), and White (7704/8392, 91.8%); had high socioeconomic status (income ≥US $100,000: 4389/8392, 52.3%; bachelor's degree or higher: 7251/8392, 86.4%); and resided in a metropolitan area core (RUCA 1: 7192/8392, 85.7%) (Table 1).

As shown in Table 2, at baseline, participants reported moderate stress (mean score 18.12, SD 6.29), where a score of 14 to 26 is considered moderate [23,24]; borderline abnormal levels of depression (mean score 8.89, SD 4.12), where a score of 8 to10 is considered borderline abnormal [20]; normal levels of anxiety (mean score 5.87, SD 3.61), where a score of 0 to 7 is considered normal [20]; and being physically active (mean 4.89 days/week, SD 2.29).

Table 2. Baseline psychological outcomes, physical activity, and perceived effects of COVID-19.

Outcome, physical activity, or effect	Mean (SD)
Psychological outcome	
Stress score[a]	18.12 (6.29)
Depression score[b]	8.89 (4.12)
Anxiety score[b]	5.87 (3.61)
Physical activity (days/week)[c]	
Pre–COVID-19 physical activity	4.87 (2.04)
Current physical activity	4.89 (2.29)
Perceived effect[d] of COVID-19 on:	
Stress	1.82 (0.80)
Mental health	2.07 (0.83)
Physical activity	2.72 (1.29)

[a]Stress was measured using the Perceived Stress Scale; summed scores range from 0 to 40, with higher scores indicating higher levels of perceived stress.

[b]Depression and anxiety were measured using the depression and anxiety subscales, respectively, of the Hospital Anxiety and Depression Scale; summed scores range from 0 to 21 (normal: 0-7; borderline abnormal: 8-10; abnormal: 11-21).

[c]Participants self-reported how many days per week (on a scale of 0 to 7) of physical activity they participated in.

[d]Participants responded to investigator-developed items; response scores range from 1 to 5, with higher scores indicating higher perceived effects of COVID-19.

As shown in Table 3, in both unadjusted and adjusted regression models, rural or urban status was not significantly associated with baseline stress, depression, anxiety, physical activity, or perceived effects of COVID-19 on stress, mental health, and physical activity. In our alternative analyses where we treated RUCA as a continuous variable, we also found no significant associations between rural or urban status and mental health outcomes, physical activity, or perceived effects of COVID-19 (Multimedia Appendix 1).

Table 3. Unadjusted and adjusted regression models exploring associations between rural or urban status and psychological outcomes, physical activity, and perceived effects of COVID-19.

Model	Rural or urban difference, β (SE)	*P* value
Unadjusted model		
Psychological outcome		
Stress	0.30 (0.22)	.09
Depression	0.15 (0.14)	.29
Anxiety	–0.15 (0.12)	.24
Physical activity		
Pre–COVID-19 physical activity	0.08 (0.07)	.27
Current physical activity	0.08 (0.08)	.31
Perceived effect of COVID-19 on:		
Stress	–0.05 (0.03)	.10
Mental health	–0.07 (0.03)	.11
Physical activity	–0.05 (0.04)	.29
Adjusted model[a]		
Psychological outcome		
Stress	0.02 (0.22)	.91
Depression	0.07 (0.15)	.64
Anxiety	–0.24 (0.13)	.07
Physical activity		
Pre–COVID-19 physical activity	0.12 (0.08)	.10
Current physical activity	0.16 (0.09)	.05
Perceived effect of COVID-19 on:		
Stress	–0.05 (0.03)	.09
Mental health	–0.07 (0.03)	.22
Physical activity	–0.02 (0.05)	.67

[a]The model controlled for gender, age, ethnicity, race, income, education, employment, Calm app use, and statewide COVID-19 prevalence at baseline.

As shown in Table 4, results from repeated-measures ANOVAs showed no significant changes in stress or perceived effects of COVID-19 on stress and mental health from baseline to follow-up. In models of depression, anxiety, and effect of COVID-19 on physical activity, there were significant main effects of time, which showed that, overall, symptoms of depression, anxiety, and perceived effect of COVID-19 on physical activity decreased from baseline to follow-up. Furthermore, there were significant main effects of age on most outcomes, which showed that, at baseline, older participants had greater stress, depression, and anxiety; engaged in more physical activity; and perceived that COVID-19 had a greater effect on their stress and mental health. There were also significant time × age interactions in models of stress and anxiety, such that older participants had smaller decreases in stress and anxiety over time than did younger participants. Results of full models with covariates can be found in Multimedia Appendix 1.

There were no significant main effects of rural or urban status, statewide COVID-19 prevalence at baseline, or change in statewide COVID-19 prevalence in any of the models (Table 4). In addition, there were no significant time × rural or urban status interactions, time × statewide COVID-19 prevalence at baseline interactions, or time × change in statewide COVID-19 prevalence interactions (Table 4). Lastly, we found no significant three-way interactions (ie, time × statewide COVID-19 prevalence at baseline × rural or urban status and time × change in statewide COVID-19 prevalence × rural or urban status).

Table 4. Results of repeated-measures analyses of variance for baseline and follow-up study outcomes[a].

Model	Baseline score, mean (SD)	Follow-up score, mean (SD)	*F* test (*df*)	*P* value
Stress	18.013 (6.174)	17.180 (6.322)		
Time			0.049 (1, 2636)	.83
Rural-urban commuting area (RUCA)			0.160 (1, 2636)	.69
Time × RUCA			0.000 (1, 2636)	.99
COVID-19 state baseline			1.078 (1, 2636)	.30
Time × COVID-19 state baseline			0.015 (1, 2636)	.90
Time × COVID-19 state baseline × RUCA			0.116 (1, 2636)	.73
COVID-19 state change			0.069 (1, 2636)	.79
Time × COVID-19 state change			0.006 (1, 2636)	.94
Time × COVID-19 state change × RUCA			0.168 (1, 2636)	.68
Depression	8.832 (4.097)	8.396 (4.156)		
Time			3.868 (1, 2613)	.049
RUCA			0.017 (1, 2613)	.90
Time × RUCA			0.339 (1, 2613)	.56
COVID-19 state baseline			0.386 (1, 2613)	.54
Time × COVID-19 state baseline			0.112 (1, 2613)	.74
Time × COVID-19 state baseline × RUCA			0.047 (1, 2613)	.83
COVID-19 state change			0.038 (1, 2613)	.85
Time × COVID-19 state change			0.519 (1, 2613)	.47
Time × COVID-19 state change × RUCA			0.183 (1, 2613)	.67
Anxiety	5.755 (3.590)	5.474 (3.663)		
Time			4.648 (1, 2613)	.03
RUCA			0.409 (1, 2613)	.52
Time × RUCA			0.353 (1, 2613)	.55
COVID-19 state baseline			0.022 (1, 2613)	.88
Time × COVID-19 state baseline			0.626 (1, 2613)	.43
Time × COVID-19 state baseline × RUCA			1.121 (1, 2613)	.29
COVID-19 state change			0.211 (1, 2613)	.65
Time × COVID-19 state change			0.010 (1, 2613)	.92
Time × COVID-19 state change × RUCA			0.310 (1, 2613)	.58
Physical activity	4.930 (2.305)	4.950 (2.243)		
Time			0.188 (1, 2519)	.67
RUCA			1.227 (1, 2519)	.27
Time × RUCA			1.248 (1, 2519)	.26
COVID-19 state baseline			0.046 (1, 2519)	.83
Time × COVID-19 state baseline			0.659 (1, 2519)	.42
Time × COVID-19 state baseline × RUCA			0.779 (1, 2519)	.38
COVID-19 state change			0.000 (1, 2519)	>.99
Time × COVID-19 state change			2.061 (1, 2519)	.15
Time × COVID-19 state change × RUCA			3.550 (1, 2519)	.06
Effect of COVID-19 on stress	1.830 (0.821)	1.790 (0.783)		

Model	Baseline score, mean (SD)	Follow-up score, mean (SD)	*F* test (*df*)	*P* value
Time			0.717 (1, 2487)	.40
RUCA			0.189 (1, 2487)	.66
Time × RUCA			3.344 (1, 2487)	.07
COVID-19 state baseline			0.935 (1, 2487)	.33
Time × COVID-19 state baseline			2.506 (1, 2487)	.11
Time × COVID-19 state baseline × RUCA			1.077 (1, 2487)	.30
COVID-19 state change			0.002 (1, 2487)	.97
Time × COVID-19 state change			1.203 (1, 2487)	.27
Time × COVID-19 state change × RUCA			0.084 (1, 2487)	.77
Effect of COVID-19 on mental health	2.110 (0.846)	2.030 (0.821)		
Time			0.081 (1, 2481)	.78
RUCA			0.138 (1, 2481)	.71
Time × RUCA			0.002 (1, 2481)	.96
COVID-19 state baseline			0.123 (1, 2481)	.73
Time × COVID-19 state baseline			0.255 (1, 2481)	.61
Time × COVID-19 state baseline × RUCA			0.510 (1, 2481)	.48
COVID-19 state change			0.069 (1, 2481)	.79
Time × COVID-19 state change			0.006 (1, 2481)	.94
Time × COVID-19 state change × RUCA			0.168 (1, 2481)	.68
Effect of COVID-19 on physical activity	2.740 (1.278)	2.670 (1.261)		
Time			3.868 (1, 2483)	.049
RUCA			0.017 (1, 2483)	.90
Time × RUCA			0.339 (1, 2483)	.56
COVID-19 state baseline			0.386 (1, 2483)	.54
Time × COVID-19 state baseline			0.112 (1, 2483)	.74
Time × COVID-19 state baseline × RUCA			0.047 (1, 2483)	.83
COVID-19 state change			0.038 (1, 2483)	.85
Time × COVID-19 state change			0.519 (1, 2483)	.47
Time × COVID-19 state change × RUCA			0.183 (1, 2483)	.67

[a]Baseline and follow-up mean (SD) values were only reported for each within-subjects factor, while *F* test (*df*) values and *P* values were only reported for each between-subjects factor and interaction effects.

Discussion

Principal Findings

In this study, we explored associations between rural or urban status, psychological outcomes, and physical activity among meditation app users. We additionally explored associations between rural or urban status and perceived effects of the ongoing COVID-19 pandemic on stress, mental health, and physical activity, as well as changes in study outcomes among rural versus urban app users over time. Overall, we found no significant associations between rural or urban status, psychological outcomes, physical activity, or perceived effects of COVID-19 on stress, mental health, and physical activity at baseline. We also found that there were significant decreases in depression, anxiety, and perceived effects of COVID-19 on physical activity over time, but there were no significant changes in stress, physical activity participation, or perceived effects of COVID-19 on stress or mental health. Additionally, there were no significant time × statewide COVID-19 prevalence × rural or urban status interactions.

Our findings generally contrast previous studies that demonstrated rural and urban differences in physical and mental health outcomes and behaviors [1-6]. Specifically, some studies found a higher prevalence of major mental disorders, including mood and anxiety disorders, in urban areas [2,25]. For example, one literature review concluded that living in urban cities was associated with a considerably higher risk for schizophrenia [2]; in addition, another meta‐analysis of urban and rural

differences in psychiatric disorders, which was conducted on data taken from 20 population survey studies, found that prevalence rates for psychiatric disorders were significantly higher in urban areas compared with rural areas [25]. Other studies found a higher prevalence of depression in rural areas [16] and have shown that rural residents receive less mental health treatment despite poorer mental health [17]. For example, a cross-sectional study using the National Health Interview Survey found that depression prevalence was significantly higher among rural populations than among urban populations [16]. Another study using data from the Medical Expenditure Panel Survey concluded that the odds of receiving any mental health treatment and specialized mental health treatment were 47% and 72% higher, respectively, for metropolitan residents compared to those living in the most rural settings [17].

Regarding health behaviors such as physical activity, rural residents have consistently been less physically active compared to their urban counterparts, which has been attributed to increased barriers and limited physical activity resources in rural areas [6,18,26]. For example, an analysis of data from the Behavioral Risk Factor Surveillance System, conducted among 398,208 adults, demonstrated that residents of nonmetropolitan counties had a lower prevalence of meeting national physical activity guidelines compared to their metropolitan counterparts [18]. These referenced studies exploring rural and urban differences in health outcomes and behaviors have generally been conducted among adults outside the context of participation in a behavioral health intervention or use of an available behavioral health program, unlike this study, which explored these differences among users of a meditation app.

This was the first study, to our knowledge, to assess rural and urban differences among meditation app users during the COVID-19 pandemic. Our results suggest that the rural residents in our sample who use meditation apps have access to tools that may address rural health disparities. For example, meditation app users need to own a mobile device and have internet access, which are social determinants of health [27]. Furthermore, mobile meditation app users likely possess digital literacy (ie, the ability to understand and utilize electronic resources or identify, access, and use electronic information from networks as well as having the skills to decipher texts, sounds, or images) [28,29]. Lastly, seeking out, downloading, and using a meditation app to manage health-related outcomes and symptoms reflect health literacy (ie, the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions) [30], which is associated with mHealth app use [31]. Overall, a combination of mobile device ownership, internet access, and digital and health literacy may be important tools for addressing rural health disparities. The use of a meditation app, specifically, may further contribute to addressing disparities via improvement of mental health outcomes and health behaviors among rural residents. Thus, research should continue to explore and establish the use of meditation apps for health promotion in both rural and urban populations.

There were no associations between rural or urban status and perceived effects of the COVID-19 pandemic on stress, mental health, or physical activity. Emerging literature has demonstrated that stress and symptoms of anxiety and depression are common psychological responses to COVID-19 [32], but studies have not yet focused on rural populations or assessed rural and urban differences in perceived effects of COVID-19. Our results suggest that during this ongoing pandemic, mobile device ownership, internet access, digital and health literacy, and meditation app use may address disparities related to the perceived effects of COVID-19 on health and health behaviors. This further reinforces the importance of using meditation apps among both rural and urban populations. However, it is important to note that our sample largely consisted of non-Hispanic White women with relatively high socioeconomic status, and may not be reflective of rural or urban mHealth app users from lower-income or racial and ethnic minority backgrounds, and the benefits of meditation apps and their potential for reducing disparities must be confirmed with studies in these underrepresented populations. Furthermore, given that the majority of our sample was urban, the absence of rural and urban differences in outcomes must be confirmed in a larger, more representative sample.

We found overall decreases in depression, anxiety, and perceived effect of COVID-19 on physical activity and no overall changes from baseline to follow-up in stress, physical activity, or perceived effects of COVID-19 on stress and mental health. Prior studies have generally demonstrated a reduction in physical activity since the start of the pandemic [33-35], and our results suggest that meditation apps may be a potential strategy for counteracting decreases in physical activity during the pandemic. Studies assessing changes in psychological outcomes over the course of the COVID-19 pandemic have resulted in mixed findings; for example, one study found increases in depression and decreases in anxiety in Argentina [36], another study found increases in anxiety in the United Kingdom [37], and another found no changes in stress, depression, or anxiety in China [38]. However, there have been limited studies longitudinally assessing changes in mental health outcomes during the pandemic in the United States. Our results provide preliminary evidence that despite the increase in state-level COVID-19 cases in the United States, there may not be negative cumulative effects of the pandemic on psychological outcomes and health behaviors among our sample (ie, mainly urban, White, female, and of high socioeconomic status) of meditation app users. However, further studies are necessary to understand longer-term effects of COVID-19 in this group, and to establish whether meditation delivered via mobile apps is an effective strategy for promoting health outcomes and behaviors over the course of the pandemic. Our results further suggest no significant effect of rural or urban status and statewide COVID-19 prevalence—both at baseline and change from baseline to follow-up—on changes in outcomes over time. However, more research is needed to elucidate time-varying changes in psychological and behavioral outcomes between rural and urban meditation app users, which could help to identify at-risk groups for interventions.

Limitations

This study was the first to explore associations between rural or urban status and psychological outcomes, physical activity, and perceived effects of COVID-19 on stress, mental health,

and physical activity among meditation app users. However, there were limitations that should be noted. Our sample was primarily female, non-Hispanic, and White and had high socioeconomic status, limiting generalizability of findings to other populations, such as racial and ethnic minorities and low-income individuals. The majority of our sample was further categorized as urban, with 85.7% of participants residing in the US Department of Agriculture's definition of a metropolitan core area (ie, RUCA 1). Although these sample demographics are frequently reflected in research assessing rural and urban differences, with studies including 80.0% to 88.7% of participants residing in RUCA 1 or the highest level of urbanization [3,39-41], future research in this area should aim to include a larger proportion of rural residents.

Furthermore, all outcomes in this study were assessed via self-report measures, which are subject to social desirability and recall bias. In addition, another limitation was the use of investigator-developed survey questions as opposed to validated questionnaires. However, for some outcomes, such as the extent to which COVID-19 impacted outcomes, there were no validated instruments when this study was conducted. With regard to behavioral outcomes (ie, physical activity and meditation), future studies in this area should aim to use device-based measures. Physical activity, in particular, is one of the outcomes to be particularly cautious about when interpreting study results. Physical activity is often overreported by participants in research studies, with respondents reporting higher rates of, or more frequent, activity than actual behavior warrants, which in turn causes self-reported measures of physical activity to suffer from low validity [42]. Thus, given our self-reported items for measuring physical activity in this study, it is important to note that these results are preliminary and must be validated in larger studies using device-based measures, which may be complemented with self-report measures to provide activity type and context. Another limitation was inherent to the study design, which included a national survey conducted among a nonrandom convenience sample. Although this study was able to contribute novel, preliminary findings, future research should assess associations between rural or urban status and health outcomes among mHealth app users in a randomized controlled trial in order to decrease bias and increase scientific rigor. Lastly, although we included total state-level COVID-19 cases over time in the analyses, a combination of additional factors, such as lockdowns, hospitalizations, and deaths, may have had an influence on participants' psychological and physical activity outcomes and perceived effects of COVID-19 on health and behaviors. Further research is also needed to establish whether cumulative effects of the pandemic will become more apparent after a longer period of time.

Conclusions

Overall, we did not find any associations between rural or urban status and psychological outcomes (ie, stress, depression, and anxiety), pre–COVID-19 and current physical activity, or perceived effects of COVID-19 on stress, mental health, and physical activity among users of a meditation app. Rural or urban status does not appear to drive differences in outcomes among meditation app users, and the use of mHealth apps should continue to be explored as a health promotion strategy in both rural and urban populations. Our results also did not show negative cumulative effects of COVID-19 on psychological outcomes and physical activity in our sample, and research should continue to explore meditation apps as a health promotion strategy during the pandemic.

Acknowledgments

The researchers would like to thank all the participants who took part in this study.

Authors' Contributions

JH and CS conceptualized the study and guided the study design. NB conducted the data analyses and drafted the manuscript. MP assisted with data analyses and interpretation of results. All authors critically reviewed and edited the manuscript and approved its submission.

Conflicts of Interest

JH is currently consulting as the Director of Science at Calm. JH has been conducting research with Calm as a partner for 6 years prior to the Director of Science role. JH directs the Scientific Advisory Board at Calm. JH has no stock in Calm and receives no financial incentives from the sales of Calm.

Multimedia Appendix 1

Unadjusted and adjusted regression models exploring associations between rural-urban status (ie, continuous rural-urban commuting area [RUCA]) and mental health outcomes, physical activity, and perceived effects of COVID-19 as well as results of repeated measures analyses.

[XLSX File (Microsoft Excel File), 30 KB - mhealth_v9i5e26037_app1.xlsx]

References

1. Henley SJ, Anderson RN, Thomas CC, Massetti GM, Peaker B, Richardson LC. Invasive cancer incidence, 2004-2013, and deaths, 2006-2015, in nonmetropolitan and metropolitan counties - United States. MMWR Surveill Summ 2017 Jul 07;66(14):1-13 [FREE Full text] [doi: 10.15585/mmwr.ss6614a1] [Medline: 28683054]

2. Gruebner O, Rapp M, Adli M, Kluge U, Galea S, Heinz A. Cities and mental health. Dtsch Arztebl Int 2017 Feb 24;114(8):121-127 [FREE Full text] [doi: 10.3238/arztebl.2017.0121] [Medline: 28302261]
3. Kozhimannil KB, Interrante JD, Henning-Smith C, Admon LK. Rural-urban differences in severe maternal morbidity and mortality in the US, 2007-15. Health Aff (Millwood) 2019 Dec;38(12):2077-2085. [doi: 10.1377/hlthaff.2019.00805] [Medline: 31794322]
4. O'Connor A, Wellenius G. Rural-urban disparities in the prevalence of diabetes and coronary heart disease. Public Health 2012 Oct;126(10):813-820. [doi: 10.1016/j.puhe.2012.05.029] [Medline: 22922043]
5. Suliga E. Nutritional behaviours of pregnant women in rural and urban environments. Ann Agric Environ Med 2015;22(3):513-517 [FREE Full text] [doi: 10.5604/12321966.1167725] [Medline: 26403126]
6. Whitfield GP, Carlson SA, Ussery EN, Fulton JE, Galuska DA, Petersen R. Trends in meeting physical activity guidelines among urban and rural dwelling adults - United States, 2008-2017. MMWR Morb Mortal Wkly Rep 2019 Jun 14;68(23):513-518 [FREE Full text] [doi: 10.15585/mmwr.mm6823a1] [Medline: 31194722]
7. Chandrashekar P. Do mental health mobile apps work: Evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
8. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: A systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
9. 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]
10. Healthcare facilities: Managing operations during the COVID-19 pandemic. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html [accessed 2020-11-05]
11. Mallow JA, Theeke LA, Barnes ER, Whetsel T, Mallow BK. Using mHealth tools to improve rural diabetes care guided by the chronic care model. Online J Rural Nurs Health Care 2014;14(1):43-65 [FREE Full text] [doi: 10.14574/ojrnhc.v14i1.276] [Medline: 26029005]
12. Yin Z, Lesser J, Paiva KA, Zapata J, Moreno-Vasquez A, Grigsby TJ, et al. Using mobile health tools to engage rural underserved individuals in a diabetes education program in South Texas: Feasibility study. JMIR Mhealth Uhealth 2020 Mar 24;8(3):e16683 [FREE Full text] [doi: 10.2196/16683] [Medline: 32207694]
13. Ostafin BD, Chawla N, Bowen S, Dillworth TM, Witkiewitz K, Marlatt GA. Intensive mindfulness training and the reduction of psychological distress: A preliminary study. Cogn Behav Pract 2006 Aug;13(3):191-197. [doi: 10.1016/j.cbpra.2005.12.001]
14. Edenfield TM, Saeed SA. An update on mindfulness meditation as a self-help treatment for anxiety and depression. Psychol Res Behav Manag 2012 Oct:131. [doi: 10.2147/prbm.s34937]
15. Bigliassi M, Bertuzzi R. Exploring the use of meditation as a valuable tool to counteract sedentariness. Front Psychol 2020;11:299 [FREE Full text] [doi: 10.3389/fpsyg.2020.00299] [Medline: 32158418]
16. Probst JC, Laditka SB, Moore CG, Harun N, Powell MP, Baxley EG. Rural-urban differences in depression prevalence: Implications for family medicine. Fam Med 2006 Oct;38(9):653-660 [FREE Full text] [Medline: 17009190]
17. Hauenstein EJ, Petterson S, Rovnyak V, Merwin E, Heise B, Wagner D. Rurality and mental health treatment. Adm Policy Ment Health 2007 May;34(3):255-267. [doi: 10.1007/s10488-006-0105-8] [Medline: 17165139]
18. Matthews KA, Croft JB, Liu Y, Lu H, Kanny D, Wheaton AG, et al. Health-related behaviors by urban-rural county classification - United States, 2013. MMWR Surveill Summ 2017 Feb 03;66(5):1-8 [FREE Full text] [doi: 10.15585/mmwr.ss6605a1] [Medline: 28151923]
19. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
21. Rural-urban commuting area codes. United States Department of Agriculture, Economic Research Service. 2020. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2020-11-06]
22. United States COVID-19 cases and deaths by state over time. Centers for Disease Control and Prevention. 2020. URL: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36 [accessed 2020-11-05]
23. Menon P, Chaudhury S, Saldanha D, Sahu S, Singh V, Pathak V. Stress levels and its association with self-harm and risk-taking behavior in medical undergraduates. Ind Psychiatry J 2018;27(1):41-46 [FREE Full text] [doi: 10.4103/ipj.ipj_31_18] [Medline: 30416290]
24. Sahu M, Gandhi S, Sharma MK, Marimuthu P. Perceived stress and resilience and their relationship with the use of mobile phone among nursing students. Invest Educ Enferm 2019 Sep;37(3):e05 [FREE Full text] [doi: 10.17533/udea.iee.v37n3e05] [Medline: 31830403]
25. Peen J, Schoevers RA, Beekman AT, Dekker J. The current status of urban-rural differences in psychiatric disorders. Acta Psychiatr Scand 2010 Feb;121(2):84-93. [doi: 10.1111/j.1600-0447.2009.01438.x] [Medline: 19624573]

26. Gilbert AS, Duncan DD, Beck AM, Eyler AA, Brownson RC. A qualitative study identifying barriers and facilitators of physical activity in rural communities. J Environ Public Health 2019;2019:7298692 [FREE Full text] [doi: 10.1155/2019/7298692] [Medline: 31341485]
27. Daniel H, Bornstein SS, Kane GC. Addressing social determinants to improve patient care and promote health equity: An American College of Physicians position paper. Ann Intern Med 2018 Apr 17;168(8):577. [doi: 10.7326/m17-2441]
28. Bawden D. Information and digital literacies: A review of concepts. J Doc 2001 Apr;57(2):218-259. [doi: 10.1108/eum0000000007083]
29. Noh Y. A study on the effect of digital literacy on information use behavior. J Libr Inf Sci 2016 Jul 09;49(1):26-56. [doi: 10.1177/0961000615624527]
30. Berkman ND, Davis TC, McCormack L. Health literacy: What is it? J Health Commun 2010;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
31. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: The potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
32. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
33. Duncan GE, Avery AR, Seto E, Tsang S. Perceived change in physical activity levels and mental health during COVID-19: Findings among adult twin pairs. PLoS One 2020;15(8):e0237695 [FREE Full text] [doi: 10.1371/journal.pone.0237695] [Medline: 32790745]
34. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide effect of COVID-19 on physical activity: A descriptive study. Ann Intern Med 2020 Nov 03;173(9):767-770. [doi: 10.7326/m20-2665]
35. Ingram J, Maciejewski G, Hand CJ. Changes in diet, sleep, and physical activity are associated with differences in negative mood during COVID-19 lockdown. Front Psychol 2020;11:588604 [FREE Full text] [doi: 10.3389/fpsyg.2020.588604] [Medline: 32982903]
36. Canet-Juric L, Andrés ML, Del Valle M, López-Morales H, Poó F, Galli JI, et al. A longitudinal study on the emotional impact cause by the COVID-19 pandemic quarantine on general population. Front Psychol 2020;11:565688 [FREE Full text] [doi: 10.3389/fpsyg.2020.565688] [Medline: 33071893]
37. Kwong A, Pearson R, Adams M, Northstone K, Tilling K, Smith D, et al. Mental health before and during the COVID-19 pandemic in two longitudinal UK population cohorts. Br J Psychiatry 2020 Nov 24:1-10 [FREE Full text] [doi: 10.1192/bjp.2020.242] [Medline: 33228822]
38. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
39. Euler R, Jimenez EY, Sanders S, Kuhlemeier A, Van Horn ML, Cohen D, et al. Rural-urban differences in baseline dietary intake and physical activity levels of adolescents. Prev Chronic Dis 2019 Jan 03;16:E01 [FREE Full text] [doi: 10.5888/pcd16.180200] [Medline: 30605423]
40. Segel JE, Lengerich EJ. Rural-urban differences in the association between individual, facility, and clinical characteristics and travel time for cancer treatment. BMC Public Health 2020 Feb 06;20(1):196 [FREE Full text] [doi: 10.1186/s12889-020-8282-z] [Medline: 32028942]
41. Margerison CE, Goldman-Mellor S. Association between rural residence and nonfatal suicidal behavior among California adults: A population-based study. J Rural Health 2019 Mar;35(2):262-269 [FREE Full text] [doi: 10.1111/jrh.12352] [Medline: 30703850]
42. Brenner PS, DeLamater JD. Social desirability bias in self-reports of physical activity: Is an exercise identity the culprit? Soc Indic Res 2013 Jun 1;117(2):489-504. [doi: 10.1007/s11205-013-0359-y]

Abbreviations

ANOVA: analysis of variance
HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression Scale–anxiety subscale
HADS-D: Hospital Anxiety and Depression Scale–depression subscale
mHealth: mobile health
RUCA: rural-urban commuting area
SQL: Structured Query Language

Edited by L Buis; submitted 25.11.20; peer-reviewed by J Brewer, AS Aslam, D Szinay, C Emezue; comments to author 08.01.21; revised version received 27.02.21; accepted 20.04.21; published 26.05.21.

Please cite as:
Bhuiyan N, Puzia M, Stecher C, Huberty J
Associations Between Rural or Urban Status, Health Outcomes and Behaviors, and COVID-19 Perceptions Among Meditation App Users: Longitudinal Survey Study
JMIR Mhealth Uhealth 2021;9(5):e26037
URL: https://mhealth.jmir.org/2021/5/e26037
doi:10.2196/26037
PMID:33900930

Original Paper

Examining the Impact of a Mobile Health App on Functional Movement and Physical Fitness: Pilot Pragmatic Randomized Controlled Trial

Matthew Jordan Stork[1], PhD; Ethan Gordon Bell[1], BHK; Mary Elizabeth Jung[1], PhD

School of Health and Exercise Sciences, The University of British Columbia, Kelowna, BC, Canada

Corresponding Author:
Mary Elizabeth Jung, PhD
School of Health and Exercise Sciences
The University of British Columbia
3333 University Way
Kelowna, BC, V1V 1VY
Canada
Phone: 1 250 807 9670
Email: mary.jung@ubc.ca

Abstract

Background: Numerous mobile apps available for download are geared toward health and fitness; however, limited research has evaluated the real-world effectiveness of such apps. The movr app is a mobile health app designed to enhance physical functioning by prescribing functional movement training based on individualized movement assessments. The influence of the movr app on functional movement and physical fitness (flexibility, strength, and cardiovascular fitness) has not yet been established empirically.

Objective: This study aims to examine the real-world impact of the movr app on functional movement, flexibility, strength, and cardiovascular fitness.

Methods: A total of 48 healthy adults (24 women and 24 men; mean age 24, SD 5 years) completed an 8-week pilot pragmatic randomized controlled trial in which they were randomly assigned to either 8-week use of the movr app (n=24) or 8-week waitlist control (n=24). Measures of functional movement (Functional Movement Screen [FMS]), strength (push-ups, handgrip strength, and countermovement jump), flexibility (shoulder flexibility, sit and reach, active straight leg raise [ASLR], and half-kneeling dorsiflexion), and cardiovascular fitness (maximal oxygen uptake [ ]) were collected at baseline and the 8-week follow-up.

Results: Repeated measures analyses of variance revealed significant group-by-time interactions for the 100-point FMS (P<.001), shoulder flexibility (P=.01), ASLR (P=.001), half-kneeling dorsiflexion (P<.001), and push-up tests (P=.03). Pairwise comparisons showed that FMS scores increased from pre- to postintervention for those in the movr group (P<.001) and significantly decreased for those in the control group (P=.04). For shoulder flexibility, ASLR, half-kneeling dorsiflexion, and push-up tests, improvements from pre- to postintervention were found in the movr group (all values of P<.05) but not in the control group (all values of P>.05). There were no changes in the sit and reach or handgrip strength test scores for either group (all values of P>.05). A significant main effect of time was found for the countermovement jump (P=.02), such that scores decreased from pre- to postintervention in the control group (P=.02) but not in the movr group (P=.38). Finally, a significant group-by-time interaction was found for (P=.001), revealing that scores decreased pre- to postintervention in the control group (P<.001), but not in the movr group (P=.54).

Conclusions: The findings revealed that movr improved indices of functional movement (FMS), flexibility (shoulder, ASLR, and dorsiflexion), and muscular endurance (push-ups) over an 8-week period compared with the control group while maintaining handgrip strength, lower body power (countermovement jump), and cardiovascular fitness ( ). Thus, this study provides initial evidence of the effectiveness of the movr app for enhancing functional movement and physical fitness among healthy adults.

Trial Registration: ClinicalTrials.gov NCT04865666; https://clinicaltrials.gov/ct2/show/NCT04865666

(JMIR Mhealth Uhealth 2021;9(5):e24076) doi:10.2196/24076

KEYWORDS

mHealth; functional movement; flexibility; strength; cardiovascular fitness

Introduction

Background

Numerous mobile apps are available for download that are geared toward health and fitness [1]. Unfortunately, most of these apps are not evidence based and have been deemed to be ineffective [1-4]. Furthermore, limited research evaluating the real-world effectiveness of such apps exists [4]. In a review of the literature on mobile apps used in health interventions, the authors concluded that "the potential for scalable behavioral interventions through these technologies is promising, but largely untapped...researchers should focus on conducting rigorous RCT (randomized controlled trial) studies with adequately powered sample sizes to determine the utility of app-based health interventions" [1]. Others have echoed this notion by highlighting that although technology-based interventions (eg, those that use mobile technology) have the advantage of being cost-effective, convenient, and accessible, there is a need for RCTs that examine the efficacy of such interventions [5].

Although many health and fitness apps are designed to promote increased physical activity or exercise participation [6], very few apps are designed to enhance the quality of functional movement and physical fitness while catering to individual needs. movr is a mobile health (mHealth) app that takes a personalized approach to improve user flexibility, strength, and overall fitness by prescribing functional movement training based on a user's own movement assessments. This is meaningful because physical fitness components of strength, flexibility, and stability have been linked to health, musculoskeletal injury risk, injury treatment, and performance of activities of daily living [7]. Furthermore, improving people's physical fitness and the quality of their functional movement may subsequently help facilitate further physical activity behavior. Physical fitness and physical activity are both important for health outcomes [8], and improved fitness has been associated with improved quality of life [9].

movr uses self-reported movement assessments to identify movement deficiencies and prescribes exercises to target those deficiencies. Other user information, such as time and equipment available and desired exercise focus, is also accounted for when generating personalized workout sessions. However, the real-world impact of movr has not yet been evaluated. Before the broad uptake of this newly developed app is encouraged, it is imperative to test its utility and, in particular, its potential to influence functional movement and physical fitness.

The Functional Movement Screen (FMS) is a cost-effective tool developed to assess functional movement based on seven fundamental movement patterns [10]. A large body of research literature has investigated the use of the FMS for predicting the future risk of injury (eg, [11,12]). Given the movr app's goal of enhancing people's quality of movement, the FMS is a measurement tool that is well suited to test its effectiveness in improving individual functional movement capacity. A limitation of the FMS is that it is typically scored on a 0-3 scale (21-point system), and it has been suggested that this scaling system may not be overly sensitive to changes over time following several weeks of movement-based training (eg, [13,14]). For example, Bodden et al [15] highlighted that based on the 0-3 scoring criteria, there were broad ranges of movement quality even when participants were assigned the same score of 2 on a given movement (ie, some scores of 2 had better movement quality than others). Notwithstanding, Butler et al [13] introduced a more detailed criterion that can be reliably used to assess the FMS movements using a 100-point scale system, and Frost et al [14] have since adapted further subcriteria to the 100-point scale and have considered it to be a *research standard* version of the FMS. To our knowledge, the FMS and the physical fitness assessments administered in this study have never been used to examine the effectiveness of a mobile app for improving physical functioning.

Owing to the COVID-19 global pandemic, there has been a worldwide risk of people spending more time being sedentary in their homes and less time being physically active [16,17]. Unfortunately, these trends may persist in the foreseeable future. mHealth apps such as movr offer remote options for people to assess their functional movement in real time, work on their physical fitness, and move more frequently without the need for extensive exercise resources. Thus, there is potential for movr to not only enhance functional movement and physical fitness but also to do so in a digital format that is accessible in a time of physical and social distancing.

Purpose and Hypotheses

This study aims to examine the real-world impact of movr on functional movement, flexibility, strength, and cardiovascular fitness (maximal oxygen uptake [ ]) in a sample of healthy women and men. Given the aims of movr, it was hypothesized that from baseline to 8-week follow-up, participants in the movr group would experience improvements in functional movement, flexibility, and muscular endurance (push-ups) compared with those in the waitlist control group (H_1). Considering the movr app's primary focus on functional movement and limited focus on building aerobic fitness, grip strength, or explosive movement, it was hypothesized that , handgrip strength, and lower body power (vertical jump) would be maintained (no increase or decrease) over the 8-week intervention period for both the movr and control groups (H_2). To our knowledge, this is the first RCT study to evaluate the effects of an mHealth app on multiple indices of physical fitness and functioning, including functional movement screening.

Methods

Study Design Overview

This study consisted of baseline testing (visit 1) and an 8-week pilot pragmatic RCT [18]. This was designed with a pragmatic intent, as it was conducted in a real-world context under the

usual circumstances [19,20]. Participants were randomly assigned to either an 8-week use of the movr app or 8-week waitlist control. Randomization was stratified by self-identified gender and was completed using a random number generator. Following the 8-week intervention, participants returned to the laboratory for follow-up testing (visit 2). The first wave of 24 participants (12 movr and 12 control) completed visit 1 of the study in September and October 2019 and visit 2 in November and December 2019, whereas the second wave of 24 participants (12 movr and 12 control) completed visit 1 in November and December 2019 and visit 2 in January and February 2020.

Participants

Considering the pilot objectives of this pragmatic RCT [21] and following recommendations for calculating the sample size for such designs [22,23], a sample of 40 participants (n=20 per group) was sought. Accounting for a 20% loss to follow-up, we recruited a sample size of 50 participants (n=25 per group). Eligible participants were healthy men and women aged between 18 and 50 years who had the ability to read and write English and owned a mobile phone that could download apps from the Apple App Store or Google Play Store. Participants were excluded from the study if they had previously used the movr app or had any contraindications to exercise based on the Get Active Questionnaire. The University of British Columbia Clinical Research Ethics Board approved the study protocol, and participants were recruited through word of mouth (ie, members of the research team and laboratory group provided general study information to individuals who may have been interested in participating) and study advertisements (ie, poster advertisements on campus, sign-up sheets in classes, and via social media outlets such as Facebook). All participants provided written informed consent and received a Can $50 (US $37.79) gift card on completion of the study.

Functional Movement

The 100-point version of the FMS [13,14] was used to identify individual movement patterns at baseline and to detect any changes in mobility and stability following the 8-week intervention. The FMS is a valid and reliable screening tool for assessing whole body movement patterns [24,25] and consists of seven core movement tests (deep squat, hurdle step, inline lunge, active straight leg raise [ASLR], shoulder flexibility, trunk stability push-up, and quadruped rotary stability). The seven FMS movement screens were performed by each participant and recorded using a portable observation laboratory (Noldus) with two video cameras (one recording in the sagittal plane [side] view and the other in the frontal plane [front or back] view). The standardized FMS verbal instructions were provided by one researcher, whereas the other researcher ensured that the camera angles captured all movements. The video-based testing was completed according to the same protocols as Butler et al [13] to minimize participant burden during laboratory visits and to blind participants to their FMS scores for each of the movements (ie, the researchers did not score the movements with participants present). Scoring of the FMS was completed at a later time using the video recordings.

The first (MJS) and second (EGB) authors scored the FMS tests of the first 10 participants independently and then reached consensus using the 100-point criteria provided by Butler et al [13] and the research standard version adapted by Frost et al [14]. This process of FMS scoring of the first 10 participants was used to help clarify the use of the 100-point scoring criteria and inform the scoring of all remaining participants. EGB provided the initial scoring of all remaining participants, which was then verified by MJS. Any discrepancies were addressed, and full consensus was achieved between both authors on the final scoring.

Physical Fitness

Multimedia Appendix 1 provides full details of the procedures followed for the flexibility, muscular endurance, handgrip strength, and lower body power tests, and Multimedia Appendix 2 provides sample images of such tests.

Flexibility Tests

The shoulder reach flexibility test was used to assess upper body flexibility, and the sit and reach, ASLR, and the half-kneeling dorsiflexion tests were used to assess lower body flexibility [26,27]. For the shoulder reach test, a soft tape measure was used to measure the distance between the participants' closed fists, and for the ASLR, the range of motion in degrees was measured on each leg using a digital inclinometer (Metriks). For each of the flexibility tests (excluding sit and reach), scores from the left and right sides (arm or leg) were summed, and the total scores were reported. A lower measurement score on the shoulder flexibility test indicates greater flexibility, whereas higher scores on the sit and reach, ASLR, and half-kneeling dorsiflexion tests indicate greater flexibility.

Muscular Endurance

A push-up test was used to determine the maximal number of successive push-up repetitions that participants could complete until failure. The test protocols followed the Canadian Society for Exercise Physiology recommendations [28], including the use of a modified push-up protocol for all women in the study.

Handgrip Strength

Handgrip strength [29] was assessed using a maximal voluntary contraction of an isometric handgrip squeeze using a Smedley spring handgrip dynamometer (BASELINE). Scores from the left and right hands were summed, and the total scores were reported.

Lower Body Power

A countermovement jump was conducted using the My Jump 2 mobile app on an iPhone 8 to assess lower body power. The validity and reliability of the My Jump 2 app has been established previously [30].

Cardiovascular Fitness

Participants performed an incremental test on a cycle ergometer (Lode Excalibur Sport) as described previously (eg, Gillen et al [31]). The resistance on the cycle ergometer was automatically increased (1 W every 3 s for women and 1 W every 2 s for men) until participants reached volitional exhaustion or could no longer maintain a pedal cadence of at least 50 rpm. A metabolic cart with an automated gas collection system (Parvo Medics, TrueOne 2400) was used to continuously

collect expired gas samples, and was calculated using the mean of the highest average oxygen consumption over a 30-second period (in mL/kg/min). Peak power output in watts and maximal heart rate were also measured during each test.

Protocol

Baseline Testing (Visit 1)

Eligible participants read the consent form and provided informed consent. Participants then completed a baseline demographic questionnaire and self-reported their physical activity levels using the International Physical Activity Questionnaire—Short Form on a laboratory computer. Participants' height and body mass were then assessed using a stadiometer (Seca 700). The order of functional and fitness testing was standardized for consistency across all visits and participants and to minimize the effects of fatigue on subsequent tests. Participants started with a light warm-up consisting of 2 minutes of continuous pedaling on a cycle ergometer, followed by testing in the following order: anthropometric measures, sit and reach test, half-kneeling dorsiflexion test, handgrip strength test, countermovement jump, push-up test, seven FMS movements (including shoulder reach and ASLR tests), and test.

Before leaving the laboratory, participants in the movr group were asked to download the movr app on their mobile phones. They were then assisted in creating a unique movr account using their assigned participant ID code. The accounts were dummy accounts (with unique ID codes as names) to ensure that participants did not use their personal information. This protocol was also used to track and retrieve the participant app usage data from the movr database server.

Waitlist Control Group

Participants in the control group were instructed to maintain their usual physical activity, diet, and sleep behavior [20,32,33] for the next 8 weeks and to avoid any specialized exercise training during this period. An attention control group was avoided because of the pragmatic nature of this study and several cited issues with attention control in behavioral interventions (eg, equal attention between groups does not necessarily eliminate unintended differences between groups and attention control groups can lead to inadvertent interventions [32,33]). Following the 8-week study period, individuals in the control group were permitted to download and use the movr app if they chose to.

Movr Group

Participants in the movr group were also instructed to maintain their usual physical activity, diet, and sleep behavior for the next 8 weeks and to avoid any specialized exercise training during this period, but they were also asked to use the movr app to supplement their current activity. Within 24 hours after their first laboratory visit, participants were instructed to complete the 10 self-reported *Movement Assessment* tests through their individual movr app account. These movement tests are performed to assess mobility, motor control, and strength and are used to determine deficiencies in movement patterns and to provide prescriptive information for exercise selection within the app. Participants were prompted (via email) to complete their Movement Assessments again at 4 and 8 weeks.

The exercises prescribed through the movr app were accompanied with videos, images, and detailed instructions on how to complete them. These exercises consisted of basic movement and mobility patterns and were designed to promote functional movement. Participants were instructed to complete a total of four *Minis* and two workout *Builder* sessions per week. *Minis* consist of 5-minute sessions designed to improve participant movement capacity (ie, flexibility, motor control, and muscular strength) and are intended to be easily incorporated into an individual's daily life. There are two forms of Minis available on the app: (1) *Your Minis*, which are designed to address specific areas for improvement based on a user's most recent Movement Assessment scores and (2) *Everyday Minis*, which can be used for a variety of everyday situations (eg, taking a desk break or pre- or postexercise). Specifically, participants were asked to complete a minimum of two Your Minis out of the four total Minis to be completed per week. Workout *Builders* are designed to be longer exercise sessions tailored to the user's desired exercise time (15, 30, 45, or 60 min), equipment available (TRX band, kettlebell, chin-up bar, dumbbells, foam roller, or none), target body region (lower, upper, or whole body), and specific exercise goal (*get sweaty*, *build strength*, or *develop mobility*). Version 3.6 of the movr app was used at the start of data collection and was updated to version 4.1 over the course of the trial. No significant changes were made to the app or exercise prescription functionality throughout the updates.

Follow-up Testing (Visit 2)

Eight weeks later, participants were asked to return to the laboratory to complete follow-up self-reported physical activity and functional and fitness testing. The testing order and procedures were completed exactly as they were at visit 1.

Statistical Analyses

Separate 2 (group)×2 (time) mixed repeated measures (RM) analysis of variance (ANOVA) was conducted on the functional and physical fitness outcomes to examine between-group (movr vs control) and within-group (pre- to postintervention) differences. A mixed RM multivariate analysis of variance (MANOVA) was used for outcome variables that were conceptually intercorrelated [34], and significant *F* tests were followed by subsequent mixed RM ANOVAs. For all tests, significant effects were followed by Bonferroni-corrected pairwise comparisons to detect between- and within-group differences. The magnitude of the observed effect sizes was reported as partial eta squared (η_p^2). All analyses were conducted using SPSS version 26, and significance was set at an α level of $P<.05$.

Results

Participants

One man in the movr group did not show up for follow-up testing, and one woman in the control group dropped out of the study for reasons unrelated to the study. Thus, a total of 48

participants (24 women and 24 men) completed the study (movr=24 and control=24), and their characteristics are presented in Table 1. One woman was unable to complete the push-up test at either visit due to a chronic shoulder injury, so these data were considered as missing and were not included in the push-up test analyses.

Table 1. Participant baseline characteristics.

Variable	movr (n=24)	Control (n=24)
Age (years), mean (SD)	22.9 (5.3)	24.3 (5.3)
Body mass (kg), mean (SD)	71.2 (14.3)	70.5 (11.0)
Height (cm), mean (SD)	171.4 (11.6)	171.2 (9.3)
BMI (kg/m^2), mean (SD)	24.1 (3.5)	23.9 (2.0)
Waist circumference (cm), mean (SD)	77.9 (9.2)	78.5 (7.1)
Maximal heart rate (bpm), mean (SD)	187.6 (8.7)	187.5 (12.0)
Peak power output (W), mean (SD)	270.9 (83.5)	285.0 (81.9)
Maximal oxygen uptake (mL/kg/min), mean (SD)	40.4 (9.5)	43.1 (9.3)
Moderate- to vigorous-intensity physical activity (metabolic equivalents min per week), mean (SD)	2520.7 (1742.6)[a]	2494.2 (1417.5)
Gender, n (%)		
Women	12 (50)	12 (50)
Men	12 (50)	12 (50)
Sex, n (%)		
Female	12 (50)	12 (50)
Male	12 (50)	12 (50)
Race, n (%)		
White	18 (75)	22 (92)
Indigenous	0 (0)	1 (4)
Chinese	1 (4)	0 (0)
Southeast Asian	2 (8)	0 (0)
South Asian	1 (4)	0 (0)
Latin American	1 (4)	1 (4)
West Indian	1 (4)	0 (0)
Highest education , n (%)		
High school	11 (46)	12 (50)
Trades certificate or diploma	1 (4)	2 (8)
Nonuniversity certificate or diploma	3 (13)	0 (0)
University certificate or diploma	7 (29)	4 (17)
Postgraduate degree	2 (8)	6 (25)
Occupation, n (%)		
Working full time	1 (4)	3 (13)
Working part time	3 (13)	2 (8)
Student	18 (75)	19 (79)
Retired	1 (4)	0 (0)
Annual household income (Can $) , n (%)		
0-25,000 (US $0-US $18,895)	13 (54)	7 (29)
25,000-50,000 (US $18,895-US $37,790)	4 (17)	5 (21)
50,000-75,000 (US $37,790-US $56,685)	2 (8)	4 (17)
75,000-100,000 (US $56,685-US $75,580)	1 (4)	0 (0)
>100,000 (>US $75,580)	2 (8)	3 (13)
Prefer not to answer	2 (8)	5 (21)

[a]n=23 for moderate- to vigorous-intensity physical activity (metabolic equivalents min per week) in the movr group only because of missing data.

Functional Movement

A 2×2 mixed RM ANOVA on 100-point FMS scores showed a significant group-by-time interaction ($F_{1,46}$=47.55; P<.001; η_p^2=0.51). Pairwise comparisons revealed that FMS scores significantly increased from pre- to postintervention for those in the movr group (mean 56.08, SD 11.40 to mean 62.71, SD 10.56; P<.001) and significantly decreased for those in the control group (mean 60.60, SD 13.53, to mean 58.77, SD 13.81; P=.04; Figure 1). A breakdown of the mean scores for each of the seven different FMS movements is presented in Table 2.

Figure 1. Measurements of (A) 100-point Functional Movement Screen, (B) shoulder reach, (C) active straight leg raise, and (D) half-kneeling dorsiflexion before (pre) and after (post) the 8-week intervention period. Circles with black connecting lines represent sample means with SD error bars, whereas gray lines represent individual participant data points. Asterisks indicate significant differences between pre- and postintervention values within a given group (*P<.05; **P<.01). FMS: Functional Movement Screen.

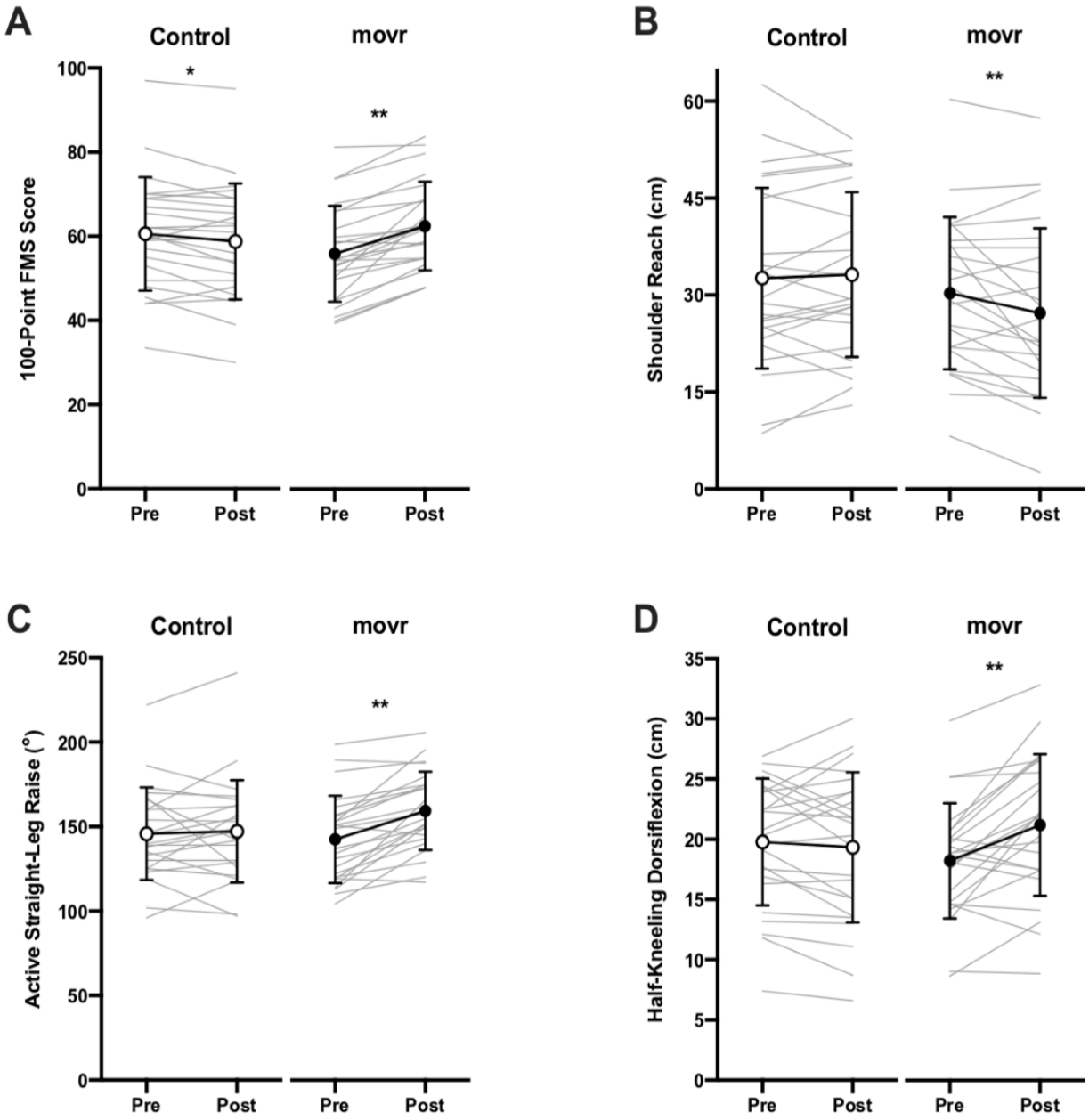


Table 2. Mean values for the 100-point Functional Movement Screen scores pre- and postintervention.

Functional Movement Screen test	movr		Control		Maximum possible test score
	Preintervention (n=24), mean (SD)	Postintervention (n=24), mean (SD)	Preintervention (n=24), mean (SD)	Postintervention (n=24), mean (SD)	
Deep squat	7.00 (3.84)	9.00 (4.22)	7.83 (4.57)	7.96 (4.41)	18
Hurdle step	14.38 (1.79)	14.67 (1.94)	14.98 (1.58)	14.35 (1.56)	18
Inline lunge	12.25 (3.67)	14.21 (3.01)	13.71 (3.28)	12.96 (3.75)	20
Shoulder mobility	6.67 (1.63)	6.92 (1.77)	6.33 (2.01)	6.17 (2.28)	8
Active straight leg raise	6.75 (3.80)	7.92 (3.36)	6.08 (4.06)	5.92 (4.31)	12
Trunk stability push-up	5.00 (4.03)	5.29 (4.24)	6.83 (4.60)	6.67 (4.56)	12
Rotary stability	4.04 (1.12)	4.71 (1.73)	4.83 (1.95)	4.75 (1.85)	12
Composite score[a]	*56.08 (11.40)*	*62.71 (10.56)*[b]	*60.60 (13.53)*	*58.77 (13.81)*[c]	*100*

[a]Italicization is to indicate that it is the total of all previous rows.

[b]Significant differences between pre- and postintervention composite Functional Movement Screen scores within a given group (P<.01).

[c]Significant differences between pre- and postintervention composite Functional Movement Screen scores within a given group (P<.05).

Physical Fitness

Flexibility Tests

A 2×2 RM MANOVA was computed across the four flexibility measures of shoulder flexibility, sit and reach, ASLR, and half-kneeling dorsiflexion. Using Pillai trace, we found that there was a significant omnibus group-by-time interaction (V=0.37; $F_{4,43}$=6.38; P<.001; η_p^2=0.37).

Shoulder Flexibility

A 2×2 mixed RM ANOVA on the shoulder reach test showed a significant group-by-time interaction ($F_{1,46}$=6.58; P=.01; η_p^2=0.13). Pairwise comparisons revealed that shoulder flexibility significantly improved (scores lowered) from pre- to postintervention for those in the movr group (mean 30.34, SD 11.80 cm, to mean 27.28, SD 13.15 cm; P=.003) but not for those in the control group (mean 32.63, SD 13.99 cm, to mean 33.16, SD 12.74 cm; P=.59; Figure 1).

Sit and Reach Test

A 2×2 mixed RM ANOVA on the sit and reach test showed no significant main effects or interaction effects (all values of P>.05).

ASLR Test

A 2×2 mixed RM ANOVA on the ASLR test showed a significant group-by-time interaction ($F_{1,46}$=11.95; P=.001; η_p^2=0.21). Pairwise comparisons revealed that ASLR significantly improved (scores increased) from pre- to postintervention for those in the movr group (mean 143.46, SD 26.03 degrees, to mean 160.50, SD 23.36 degrees; P<.001) but not for those in the control group (mean 145.79, SD 27.39 degrees, to mean 147.13, SD 30.78 degrees; P=.68; Figure 1).

Half-Kneeling Dorsiflexion

A 2×2 mixed RM ANOVA on half-kneeling dorsiflexion showed a significant group-by-time interaction ($F_{1,46}$=14.23; P<.001; η_p^2=0.24). Pairwise comparisons revealed that dorsiflexion significantly improved (scores increased) from pre- to postintervention for those in the movr group (mean 18.33, SD 4.81 cm, to mean 21.30, SD 5.91 cm; P<.001) but not for those in the control group (mean 19.77, SD 5.27 cm, to mean 19.32, SD 6.25 cm; P=.49; Figure 1).

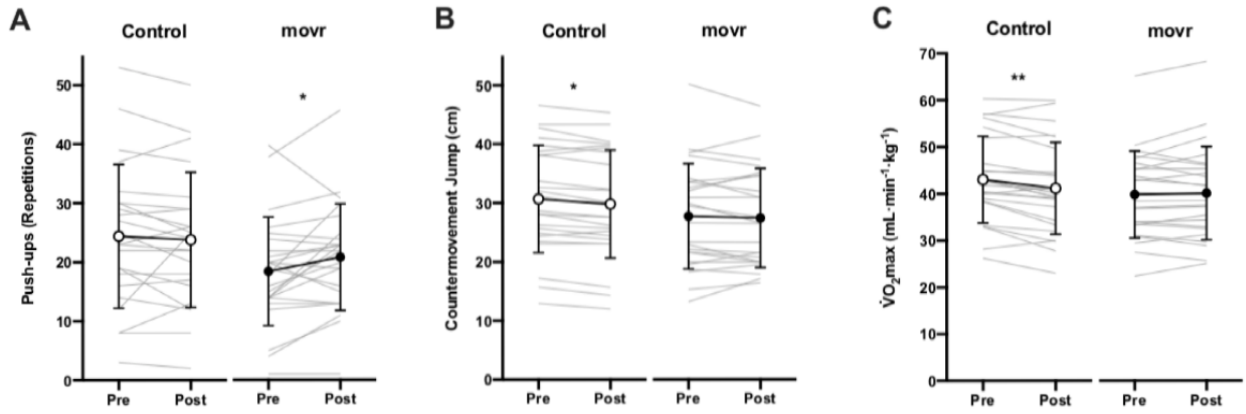
Strength and Power Tests

A 2×2 RM MANOVA was computed across the three strength and power measures of push-ups, handgrip strength, and countermovement jump. Using Pillai trace, we found that there was a significant omnibus effect of time (V=0.19; $F_{3,43}$=3.28; P=.03; η_p^2=0.19).

Push-up Test

A 2×2 mixed RM ANOVA on push-ups showed a significant group-by-time interaction ($F_{1,45}$=5.06; P=.03; η_p^2=0.10). Pairwise comparisons revealed that push-ups significantly increased from pre- to postintervention for those in the movr group (mean 18.50, SD 9.27 repetitions, to mean 20.92, SD 9.09 repetitions; P=.01) but not for those in the control group (mean 24.39, SD 12.16 repetitions, to mean 23.78, SD 11.46 repetitions; P=.53; Figure 2).

Figure 2. Measurements of (A) push-ups, (B) countermovement jump, and (C) maximal oxygen uptake (<inline-graphic xlink:href="mhealth_v9i5e24076_fig3.png" xlink:type="simple" mimetype="image"/>) before (pre) and after (post) the 8-week intervention period. Circles with black connecting lines represent sample means with SD error bars, whereas gray lines represent individual participant data points. Asterisks indicate significant differences between pre- and postintervention values within a given group (*P<.05; **P<.01). <inline-graphic xlink:href="mhealth_v9i5e24076_fig3.png" xlink:type="simple" mimetype="image"/>: maximal oxygen uptake.



Handgrip Strength

A 2×2 mixed RM ANOVA on handgrip strength showed no significant main effects or interaction effects (all values of P>.05).

Countermovement Jump

A 2×2 mixed RM ANOVA on countermovement jump showed only a significant main effect of time ($F_{1,46}$=5.79; P=.02; η_p^2=0.11). None of the other main effects or interactions were significant (all values of P>.05). Pairwise comparisons for time revealed that countermovement jump significantly decreased from pre- to postintervention for those in the control group (mean 30.69, SD 9.11 cm, to mean 29.83, SD 9.19 cm; P=.02) but not for those in the movr group (mean 27.82, SD 8.97 cm, to mean 27.52, SD 8.43 cm; P=.38; Figure 2).

Cardiovascular Fitness

A 2×2 mixed RM ANOVA on [illegible] showed a significant group-by-time interaction ($F_{1,46}$=11.50; P=.001; η_p^2=0.20). Pairwise comparisons revealed that [illegible] significantly decreased from pre- to postintervention for those in the control group (mean 43.06, SD 9.27 mL/kg/min, to mean 41.16, SD 9.84 mL/kg/min; P<.001) but not for those in the movr group (mean 40.42, SD 9.46 mL/kg/min, to mean 40.70, SD 10.14 mL/kg/min; P=.54; Figure 2).

Movr Usage Data

One woman in the movr group experienced technical issues related to the app that impacted her usage data being recorded on the movr database server. As such, her data were considered as missing and were not included in the movr usage reports. Over the 8-week study period, participants in the movr group completed an average of 2.20 (SD 1.27) Minis (mean 1.40, SD 0.83 Your Minis, and mean 0.80, SD 0.77 Everyday Minis) and an average of 0.96 (SD 0.72) Builders per week (n=23). In terms of total time spent completing sessions, participants in the movr group completed an average of 11.01 (SD 6.35) minutes per week completing Minis and an average of 17.93 (SD 14.68) minutes per week completing Builders, with a total average usage of 28.94 (SD 18.13) minutes per week (n=23). Note that the time spent completing Minis and Builders does not include the time participants spent completing the Movement Assessments.

Discussion

Principal Findings

The purpose of this study was to examine the real-world impact of movr on functional movement, strength, flexibility, and cardiovascular fitness. The main findings were that 8 weeks of using movr for an average of 29 minutes per week improved functional movement (FMS), most measures of flexibility (shoulder, ASLR, and dorsiflexion), and muscular endurance (push-ups) and led to the maintenance of handgrip strength, lower body power (countermovement jump), and cardiovascular fitness ([illegible]). These findings illustrate the potential real-world effectiveness of the movr app for enhancing physical functioning.

Functional Movement

Consistent with H_1, the 100-point FMS scores increased from pre- to postintervention for participants in the movr group compared with the control group. Interestingly, the FMS scores decreased over time for participants in the control group. Previous studies have also reported increases in FMS scores over the course of several weeks of specialized functional training or yoga [15,35,36]. In these studies, participant samples consisted of firefighters [35], American football players [36], and mixed martial arts athletes [15], and the duration of the interventions ranged from 6 to 8 weeks. Another study [14] failed to detect any changes in FMS scores following a 12-week functional training intervention among firefighters.

Notably, two of the abovementioned studies [35,36] did not include a control group in their design, three of these studies consisted of highly specialized and individualized training programs (including components that specifically targeted movement deficits identified by baseline FMS scores)

[14,15,36], and all four studies included supervised training sessions that were led by trained exercise professionals (eg, strength and conditioning specialists). In addition, the participants in each of these studies were all highly trained individuals. We are unaware of any previous studies that demonstrated improvements in FMS scoring from pre- to postintervention that included training programs that were delivered entirely remotely and unsupervised or consisted of training programs that were not prescribed by exercise specialists. As such, the current findings demonstrate the potential for mHealth apps such as movr to enhance functional movement without requiring the supervision of professionals and among individuals who are not highly trained.

FMS has become a popular screening tool among researchers and exercise practitioners (eg, strength and conditioning coaches and physical therapists) and has been used to prescribe exercise and rehabilitation programs based on identified movement deficiencies (eg, dos Santos et al [11]). There has also been continued interest in using the FMS as an injury prevention tool that can be used to detect predisposition to injury and/or predict future injury [11,12,24]. In this sense, it is possible that the improvements in FMS scores found for the movr group may have meaningful implications for reducing the risk of injury. However, this postulation should be interpreted cautiously as findings from several systematic reviews and meta-analyses have drawn conflicting conclusions. Some have supported the predictive utility of the FMS for future injury (eg, dos Santos et al [11] and Bonazza et al [24]), whereas others do not support the use of FMS as an injury prediction tool (eg, Moran et al [12]) or suggest that the heterogeneity in studied sample populations makes it challenging to draw definitive conclusions [37]. Future research on this topic is required before the injury-related implications of the current FMS findings can be considered further.

Flexibility Tests

In line with H_1, measurements of shoulder flexibility, ASLR, and half-kneeling dorsiflexion showed an improvement from pre- to postintervention for participants in the movr group but not for participants in the control group. However, inconsistent with H_1, sit and reach did not change over the course of the intervention period for either group. These findings suggest that the movr app was indeed effective for improving indices of upper and lower body flexibility, with the exception of the sit and reach test. Both shoulder flexibility and lower body flexibility are key areas that are identified via the Movement Assessments and targeted through the Minis and Builders sessions within the movr app.

Intriguingly, although the ASLR and sit and reach tests both measure components of hamstring flexibility, only ASLR showed a significant improvement for those who used movr. This may be explained, in part, by evidence that sit and reach test scores are strongly influenced by factors other than hamstring extensibility [38], such as pelvic tilt and lumbar spine flexion [39]. The sit and reach consists of passive hamstring lengthening, whereas the ASLR requires an individual to actively raise each leg under their control. Furthermore, abdominal wall bracing and lumbar stability are needed to minimize the risk of pain while performing the ASLR [40]. It may be that the movr app's focus on improving core activation, strength, and motor control may have differentially contributed to the improvements found for ASLR but not for sit and reach.

Strength and Power Tests

Consistent with H_1, the number of maximal push-up test repetitions increased from pre- to postintervention for participants in the movr group but not for participants in the control group. This is likely because of the movr app's focus on improving muscular strength and endurance through several core, stability, mobility, and strength exercises. Specifically, bodyweight push-up exercises were prescribed during the Builder sessions. Given that participants in the movr group were found to improve their shoulder flexibility over the intervention period, it is possible that this may have also facilitated an improvement in muscular endurance, as reflected in the push-up test.

In line with H_2, there were no changes in handgrip strength over the intervention period in either group. This finding is consistent with a previous 6-week yoga training RCT that found no changes in handgrip strength for the yoga or control group [41]. We are unaware of other studies measuring changes to handgrip strength following similar interventions.

Partially consistent with H_2, there were significant decreases in countermovement jump height for participants in the control group from pre- to postintervention, but no changes were found for participants in the movr group. It is unclear why participants in the control group experienced this decline, but it may be in part due to the temporal aspects of the study enrollment. It is possible that the change in seasons from fall (eg, September and October) to winter (eg, November and December; wave 1 participants) and the intervention period overlapping with seasonal holidays (eg, Christmas 2019; wave 2 participants) may have reduced overall physical activity patterns and led to potential gains in body mass. For instance, holidays typically represent a time of increased weight gain associated with increased food consumption and reduced exercise [42]. Although no significant changes in body mass from pre- to postintervention were detected for either group, very slight increases were observed in participants in the control group (mean 70.45 kg to 70.77 kg) but not for those in the movr group (mean 71.22 to 71.15 kg). Even slight increases in body mass may have been enough to reduce countermovement jump flight time and subsequently led to the decreases in countermovement jump that were found.

Cardiovascular Fitness

Partially consistent with H_2, there was a significant decrease in

from pre- to postintervention for participants in the control group, but no significant changes were observed for participants in the movr group. Although this was a surprising finding, it may also be explained by the aforementioned temporal factors associated with the enrollment timeline of the study. It may be that the transition from fall to winter, coupled with seasonal holidays, decreased physical activity and increased weight gain among participants. For example, outdoor forms of exercise

(eg, running and cycling) would have become less accessible during winter months. Thus, these factors may have subsequently hindered cardiovascular fitness. Interestingly, although there were no significant changes in self-reported physical activity behavior from pre- to postintervention for either group, there tended to be a decrease for those in the control group (mean 2494.17 to 2210.00 MET min per week), and an increase for those in the movr group (mean 2520.70 to 2662.73 MET min per week). Another possibility is that as cardiovascular fitness levels were relatively high for all participants at baseline, there may have been less room for improvement and a greater likelihood for a potential decline over time. In any case, it appears that the tendency for a decline in  that was apparent for those in the control group was not apparent for those randomized to the movr group. It is possible that the use of movr may have somehow mitigated these potential temporal factors that were seen for those in the control group; however, future investigation would be required to understand why this was the case.

Practical Implications

The 100-point FMS system was developed to increase the precision of scoring and subsequently lead to greater sensitivity in detecting changes in FMS in response to interventions [13]. However, it should be noted that the use of the 100-point system is more complex to implement and requires video analysis, which may take away from the simplicity and time efficiency of using the FMS as a diagnostic tool in practice [15]. For instance, in this study, each FMS performance was carefully analyzed using two camera angles and took approximately 150 hours each for the first and second authors to fully score. This scoring system may not be feasible for practice but may be feasible for research purposes and standards.

Improvements in FMS, flexibility (eg, shoulder flexibility, ASLR, and ankle dorsiflexion), and strength (eg, push-ups) have several potential benefits and implications for musculoskeletal health and injury prevention [7,11]. For example, sufficient ankle dorsiflexion is critical for regular activities of daily living, such as walking, running, and stairclimbing, and restricted ankle dorsiflexion can contribute to overuse injuries of the foot and lower limbs [27]. Although the current findings regarding improved functional movement, flexibility, and strength are promising, future research is encouraged to determine how the magnitude of change to these outcomes found in this study may translate into specific clinically meaningful outcomes.

Importantly, we avoided incorporating any behavior change counseling or techniques into the intervention component of this study to reflect the free-living experiences of using a downloaded app. Future researchers are encouraged to explore the added benefit of incorporating theory-based behavior change techniques along with the use of the movr app. Nonetheless, the results of this study demonstrated promising initial evidence of the benefits of movr on measures of physical functioning, despite the minimalistic nature of the intervention. This may suggest that individuals can still reap meaningful benefits from the app without requiring additional counseling or resources.

To our knowledge, there are limited to no studies that have investigated the real-world impact of mHealth apps designed specifically to improve functional movement. The findings from this study provide early evidence of the potential impact similar apps may have. It also acts as a reminder of the importance of enhancing physical function and movement as a whole as a precursor and/or facilitator to improve participation in and quality of physical activity. mHealth apps aiming to increase physical activity participation may benefit from incorporating additional components that focus on enhancing functional movement.

At this time, the COVID-19 pandemic likely continues to have a significant impact on sedentary behavior and physical inactivity, and it is unclear when this will subside [16,17]. As such, digital technology can play a considerable role in curbing such trends by providing access to health and fitness alternatives in the form of mHealth. The findings of this study provide initial evidence of the effectiveness of the movr app for enhancing functional movement and physical fitness in a digital format that is highly accessible and requires minimal resources. Therefore, movr may represent a viable mHealth option for individuals as they adapt to social distancing practices.

Strengths, Limitations, and Future Directions

This study had several strengths. As recommended in previous mHealth and technology literature [1,5], an RCT design that included a waitlist control group was implemented. We elected to use a modified research standard version of the 100-point FMS system in addition to a battery of other quantifiable flexibility, strength, and fitness assessments to allow for an interdisciplinary, comprehensive, and more sensitive assessment of functional movement and physical fitness. The intervention component of the study required minimal resources, was cost- and time-efficient, and required no additional attention from researchers. Similarly, the movr app is easy to use remotely (eg, at home, a hotel, or the gym), is highly personalized to individual needs, and only required an average time commitment of 28.94 minutes per week (11.01 min per week for Minis; 17.93 min per week for Builders) over the study period. Furthermore, this study used a pragmatic approach to capture real-world conditions. Taken together, these factors increase the ecological validity of this study. To our knowledge, this is the first randomized controlled study to evaluate the effects of an mHealth app on functional movement and physical fitness.

This study also has limitations that are worth noting. The duration of this RCT was only 8 weeks, which was sufficient to detect differences in patterns of change between conditions, but may not have been sufficient to see more drastic changes in some measures of physical fitness (eg, sit and reach, handgrip strength, and countermovement jump). Future RCTs could benefit from studying these effects over a longer period. Although participants in the movr group were ultimately free to use the app as little or as much as they wanted to on their own time, participants were aware that their app usage was being monitored. Thus, participants' experiences using movr over the study duration may have been different from what they would have been if they had used the app outside of the study parameters. Similar to any mHealth apps, there were technical

issues reported by participants over the study duration, such as the movr app crashing in the middle of a Minis or Builder session. In such cases, a workout would not have been registered as complete in the movr database; thus, it is possible that the movr usage data may have been underreported (ie, participants were using the app more than what was recorded on the movr database server). Although there was a range of activity status and age, most participants in this study were physically active, young, and healthy. Therefore, the results of the study may not generalize to individuals who are physically inactive, older, or living with a chronic disease. Future research is encouraged to determine if the current findings can be generalized to other sample populations.

Conclusions

Although there are countless mHealth apps, very few are designed specifically to enhance functional movement and physical fitness at the individual level. movr is a novel mHealth app that uses self-reported movement patterns to prescribe workouts that cater to individual user needs. This pilot pragmatic RCT was used to empirically evaluate the movr app's real-world impact for the first time. The findings revealed that movr improved indices of functional movement, flexibility, and muscular endurance over an 8-week period compared with the control group while maintaining handgrip strength, lower body power, and cardiovascular fitness. Taken together, this study demonstrated the potential of movr as an accessible mHealth app that may be used to enhance indices of functional movement and physical fitness.

Acknowledgments

The authors would like to acknowledge Mitacs and Lululemon Athletica for funding this project and movr for providing the necessary data and facilitating the research process. The authors would also like to thank Gregory duManoir for his guidance with functional and fitness assessment parameters, Sara McCreary for her assistance with data collection, Andrea Kasian and Makenna Moore for their help with setting up the web-based surveys and pilot testing of protocols, and the participants who volunteered their time. The authors gratefully acknowledge the Michael Smith Foundation for Health Research and Mitacs and Lululemon Athletica for their financial support to MJS through a Research Trainee Award and Mitacs Accelerate Internship, respectively.

Conflicts of Interest

The authors have no conflicts of interest to declare. However, as the authors have acknowledged earlier, MJS was partially funded through a Mitacs Accelerate Internship, a government-funded initiative that supports university (in this case, UBC) and industry (in this case, Lululemon Athletica) research partnerships for his involvement in this project as a postdoctoral fellow. In addition, movr facilitated the research process and provided access to their database servers. The involvement of Lululemon and movr did not influence any other aspects of the investigation, including the study findings or interpretation of the findings. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Multimedia Appendix 1
Fitness testing procedures.
[PDF File (Adobe PDF File), 76 KB - mhealth_v9i5e24076_app1.pdf]

Multimedia Appendix 2
Sample images of fitness tests: (A) shoulder reach flexibility, (B) sit and reach, (C) active straight-leg raise, (D) handgrip strength, (E) half-kneeling dorsiflexion, (F) push-up.
[PNG File , 4358 KB - mhealth_v9i5e24076_app2.png]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1951 KB - mhealth_v9i5e24076_app3.pdf]

References

1. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
2. Modave F, Bian J, Leavitt T, Bromwell J, Iii HC, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]

3. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):1-26 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
4. Direito A, Dale LP, 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(1):1-7 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
5. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126. [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
6. Petersen JM, Kemps E, Lewis LK, Prichard I. Psychological mechanisms underlying the relationship between commercial physical activity app use and physical activity engagement. Psychol Sport Exerc 2020 Nov;51:1-32. [doi: 10.1016/j.psychsport.2020.101719]
7. Micheo W, Baerga L, Miranda G. Basic principles regarding strength, flexibility, and stability exercises. PM R 2012 Nov 20;4(11):805-811. [doi: 10.1016/j.pmrj.2012.09.583] [Medline: 23174542]
8. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? Med Sci Sports Exerc 2001 Jun;33(6 Suppl):379-99, 419. [doi: 10.1097/00005768-200106001-00007] [Medline: 11427763]
9. Ståhle A, Mattsson E, Rydén L, Unden A, Nordlander R. Improved physical fitness and quality of life following training of elderly patients after acute coronary events. A 1 year follow-up randomized controlled study. Eur Heart J 1999 Oct;20(20):1475-1484. [doi: 10.1053/euhj.1999.1581] [Medline: 10493846]
10. Cook G, Burton L, Hoogenboom B. Pre-participation screening: the use of fundamental movements as an assessment of function - part 1. N Am J Sports Phys Ther 2006 May;1(2):62-72 [FREE Full text] [Medline: 21522216]
11. Bunn PD, Rodrigues AI, da Silva EB. The association between the functional movement screen outcome and the incidence of musculoskeletal injuries: a systematic review with meta-analysis. Phys Ther Sport 2019 Jan;35:146-158. [doi: 10.1016/j.ptsp.2018.11.011] [Medline: 30566898]
12. Moran RW, Schneiders AG, Mason J, Sullivan SJ. Do Functional Movement Screen (FMS) composite scores predict subsequent injury? A systematic review with meta-analysis. Br J Sports Med 2017 Dec 30;51(23):1661-1669. [doi: 10.1136/bjsports-2016-096938] [Medline: 28360142]
13. Butler RJ, Plisky PJ, Kiesel KB. Interrater reliability of videotaped performance on the functional movement screen using the 100-point scoring scale. Athl Train Sport Heal Care 2012 May 15;4(3):103-109. [doi: 10.3928/19425864-20110715-01]
14. Frost DM, Beach TA, Callaghan JP, McGill SM. Using the Functional Movement Screen™ to evaluate the effectiveness of training. J Strength Cond Res 2012;26(6):1620-1630. [doi: 10.1519/jsc.0b013e318234ec59]
15. Bodden JG, Needham RA, Chockalingam N. The effects of an intervention program on functional movement screen test scores in mixed martial arts athletes. J Strength Cond Res 2015;29(1):219-225. [doi: 10.1519/jsc.0b013e3182a480bf]
16. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide effect of COVID-19 on physical activity: a descriptive study. Ann Intern Med 2020 Nov 03;173(9):767-770. [doi: 10.7326/m20-2665]
17. Hall G, Laddu DR, Phillips SA, Lavie CJ, Arena R. A tale of two pandemics: how will COVID-19 and global trends in physical inactivity and sedentary behavior affect one another? Prog Cardiovasc Dis 2020 Apr 08:108-110 [FREE Full text] [doi: 10.1016/j.pcad.2020.04.005] [Medline: 32277997]
18. Petrella RJ, Gill DP, Zou G, DE Cruz A, Riggin B, Bartol C, et al. Hockey fans in training: a pilot pragmatic randomized controlled trial. Med Sci Sports Exerc 2017 Dec;49(12):2506-2516 [FREE Full text] [doi: 10.1249/MSS.0000000000001380] [Medline: 28719494]
19. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016 Aug 04;375(5):454-463. [doi: 10.1056/nejmra1510059]
20. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. Br Med J 2015 May 08;350:h2147. [doi: 10.1136/bmj.h2147] [Medline: 25956159]
21. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010 Jul 16;10(1):67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
22. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]
23. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Statist 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
24. Bonazza NA, Smuin D, Onks CA, Silvis ML, Dhawan A. Reliability, validity, and injury predictive value of the functional movement screen: a systematic review and meta-analysis. Am J Sports Med 2017 Mar 21;45(3):725-732. [doi: 10.1177/0363546516641937] [Medline: 27159297]
25. Smith CA, Chimera NJ, Wight NJ, Warren M. Interrater and intrarater reliability of the Functional Movement Screen. J Strength Cond Res 2013;27(4):982-987. [doi: 10.1519/jsc.0b013e3182606df2]
26. Wells K, Dillon E. The sit and reach—a test of back and leg flexibility. Res Q Am Assoc Health Phys Educ Recreat 2013 Dec 26;23(1):115-118. [doi: 10.1080/10671188.1952.10761965]

27. Bennell KL, Talbot RC, Wajswelner H, Techovanich W, Kelly DH, Hall AJ. Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. Aust J Physiother 1998;44(3):175-180. [doi: 10.1016/s0004-9514(14)60377-9] [Medline: 11676731]
28. Canadian Society for Exercise Physiology. CSEP Physical Activity Training for Health (CSEP-PATH). Ottawa, Ontario: Canadian Society for Exercise Physiology; 2013.
29. Roberts HC, Denison HJ, Martin HJ, Patel HP, Syddall H, Cooper C, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age Ageing 2011 Jul;40(4):423-429 [FREE Full text] [doi: 10.1093/ageing/afr051] [Medline: 21624928]
30. Balsalobre-Fernández C, Glaister M, Lockey RA. The validity and reliability of an iPhone app for measuring vertical jump performance. J Sports Sci 2015 Jan 02;33(15):1574-1579. [doi: 10.1080/02640414.2014.996184] [Medline: 25555023]
31. Gillen JB, Percival ME, Skelly LE, Martin BJ, Tan RB, Tarnopolsky MA, et al. Three minutes of all-out intermittent exercise per week increases skeletal muscle oxidative capacity and improves cardiometabolic health. PLoS One 2014;9(11):e111489 [FREE Full text] [doi: 10.1371/journal.pone.0111489] [Medline: 25365337]
32. Freedland KE, Mohr DC, Davidson KW, Schwartz JE. Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. Psychosom Med 2011 May;73(4):323-335 [FREE Full text] [doi: 10.1097/PSY.0b013e318218e1fb] [Medline: 21536837]
33. Freedland KE. Demanding attention: reconsidering the role of attention control groups in behavioral intervention research. Psychosom Med 2013 Feb;75(2):100-102. [doi: 10.1097/PSY.0b013e3182851b75] [Medline: 23362502]
34. Huberty CJ, Morris JD. Multivariate analysis versus multiple univariate analyses. Psychol Bull 1989;105(2):302-308. [doi: 10.1037/0033-2909.105.2.302]
35. Cowen VS. Functional fitness improvements after a worksite-based yoga initiative. J Bodyw Mov Ther 2010 Jan;14(1):50-54. [doi: 10.1016/j.jbmt.2009.02.006] [Medline: 20006289]
36. Kiesel K, Plisky PJ, Butler RJ. Functional movement test scores improve following a standardized off-season intervention program in professional football players. Scand J Med Sci Sports 2011 Apr;21(2):287-292. [doi: 10.1111/j.1600-0838.2009.01038.x] [Medline: 20030782]
37. Trinidad-Fernandez M, Gonzalez-Sanchez M, Cuesta-Vargas AI. Is a low Functional Movement Screen score (≤14/21) associated with injuries in sport? A systematic review and meta-analysis. BMJ Open Sport Exerc Med 2019 Sep 18;5(1):e000501 [FREE Full text] [doi: 10.1136/bmjsem-2018-000501] [Medline: 31673399]
38. Miyamoto N, Hirata K, Kimura N, Miyamoto-Mikami E. Contributions of hamstring stiffness to straight-leg-raise and sit-and-reach test scores. Int J Sports Med 2018 Feb 30;39(2):110-114. [doi: 10.1055/s-0043-117411] [Medline: 29190853]
39. Muyor JM, Zemková E, Štefániková G, Kotyra M. Concurrent validity of clinical tests for measuring hamstring flexibility in school age children. Int J Sports Med 2014 Jul 14;35(8):664-669. [doi: 10.1055/s-0033-1353217] [Medline: 24424962]
40. Liebenson C, Karpowicz AM, Brown SH, Howarth SJ, McGill SM. The active straight leg raise test and lumbar spine stability. PM R 2009 Jun 11;1(6):530-535. [doi: 10.1016/j.pmrj.2009.03.007] [Medline: 19627942]
41. Madanmohan, Mahadevan SK, Balakrishnan S, Gopalakrishnan M, Prakash ES. Effect of six weeks yoga training on weight loss following step test, respiratory pressures, handgrip strength and handgrip endurance in young healthy subjects. Indian J Physiol Pharmacol 2008;52(2):164-170. [Medline: 19130860]
42. Schoeller DA. The effect of holiday weight gain on body weight. Physiol Behav 2014 Jul;134:66-69. [doi: 10.1016/j.physbeh.2014.03.018] [Medline: 24662697]

Abbreviations

ANOVA: analysis of variance
ASLR: active straight leg raise
FMS: Functional Movement Screen
MANOVA: multivariate analysis of variance
mHealth: mobile health
RCT: randomized controlled trial
RM: repeated measures
VO_2 max: maximal oxygen uptake

Edited by L Buis; submitted 26.09.20; peer-reviewed by J Sieverdes, T Plante; comments to author 09.02.21; revised version received 14.03.21; accepted 25.03.21; published 28.05.21.

Please cite as:
Stork MJ, Bell EG, Jung ME
Examining the Impact of a Mobile Health App on Functional Movement and Physical Fitness: Pilot Pragmatic Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e24076
URL: https://mhealth.jmir.org/2021/5/e24076
doi:10.2196/24076
PMID:34047704

Original Paper

Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study

Silvia Gabrielli[1], MSc, PhD; Silvia Rizzi[1], MSc; Giulia Bassi[1], MSc; Sara Carbone[1], MSc; Rosa Maimone[1], MSc; Michele Marchesoni[1], MSc; Stefano Forti[1], MSc

Digital Health Lab, Fondazione Bruno Kessler, Trento, Italy

Corresponding Author:
Silvia Gabrielli, MSc, PhD
Digital Health Lab
Fondazione Bruno Kessler
Via Sommarive 18
Trento, 38123
Italy
Phone: 39 0461 312 47 ext 247
Email: sgabrielli@fbk.eu

Abstract

Background: University students are increasingly reporting common mental health problems, such as stress, anxiety, and depression, and they frequently face barriers to seeking psychological support because of stigma, cost, and availability of mental health services. This issue is even more critical in the challenging time of the COVID-19 pandemic. Digital mental health interventions, such as those delivered via chatbots on mobile devices, offer the potential to achieve scalability of healthy-coping interventions by lowering cost and supporting prevention.

Objective: The goal of this study was to conduct a proof-of-concept evaluation measuring the engagement and effectiveness of Atena, a psychoeducational chatbot supporting healthy coping with stress and anxiety, among a population of university students.

Methods: In a proof-of-concept study, 71 university students were recruited during the COVID-19 pandemic; 68% (48/71) were female, they were all in their first year of university, and their mean age was 20.6 years (SD 2.4). Enrolled students were asked to use the Atena psychoeducational chatbot for 4 weeks (eight sessions; two per week), which provided healthy-coping strategies based on cognitive behavioral therapy, positive psychology, and mindfulness techniques. The intervention program consisted of conversations combined with audiovisual clips delivered via the Atena chatbot. Participants were asked to complete web-based versions of the 7-item Generalized Anxiety Disorder scale (GAD-7), the 10-item Perceived Stress Scale (PSS-10), and the Five-Facet Mindfulness Questionnaire (FFMQ) at baseline and postintervention to assess effectiveness. They were also asked to complete the User Engagement Scale–Short Form at week 2 to assess engagement with the chatbot and to provide qualitative comments on their overall experience with Atena postintervention.

Results: Participants engaged with the Atena chatbot an average of 78 (SD 24.8) times over the study period. A total of 61 out of 71 (86%) participants completed the first 2 weeks of the intervention and provided data on engagement (10/71, 14% attrition). A total of 41 participants out of 71 (58%) completed the full intervention and the postintervention questionnaires (30/71, 42% attrition). Results from the completer analysis showed a significant decrease in anxiety symptoms for participants in more extreme GAD-7 score ranges (t_{39}=0.94; P=.009) and a decrease in stress symptoms as measured by the PSS-10 (t_{39}=2.00; P=.05) for all participants postintervention. Participants also improved significantly in the *describing* and *nonjudging* facets, based on their FFMQ subscale scores, and asked for some improvements in the user experience with the chatbot.

Conclusions: This study shows the benefit of deploying a digital healthy-coping intervention via a chatbot to support university students experiencing higher levels of distress. While findings collected during the COVID-19 pandemic show promise, further research is required to confirm conclusions.

(JMIR Mhealth Uhealth 2021;9(5):e27965) doi:10.2196/27965

KEYWORDS

mobile mental health; chatbots; anxiety; stress; university students; digital health; healthy-coping intervention; COVID-19

Introduction

Increased numbers of adults, particularly university students, are experiencing symptoms of stress, anxiety, and depression [1,2], which are exacerbated by the recent restrictions introduced because of the COVID-19 pandemic [3,4]. In addition, up to 74% of adults experience their first onset of a mental health diagnosis before the age of 24 years [5]. However, about three-quarters of college students who are in need of clinical services do not access them [6]; this is because they have low mental health literacy and do not recognize a need for treatment [7], but also because of the high cost of treatment, low availability, or attitudinal barriers, such as perceived stigma [8,9]. In recent years, wider access to digital technology and mobile phones has presented new opportunities for overcoming these barriers, by offering the possibility of delivering digital mental health interventions in a more scalable and convenient way [10,11].

Empirical studies on evidence-based digital interventions for mental health, including internet-based cognitive behavioral therapy (CBT), have shown that these interventions are effective, feasible, and acceptable to users [12-14], although some limitations have been found, mainly regarding low engagement by users and low completion rates [15,16]. The integration of human coaching and support in digital mental health interventions can help improve adherence and behavior change outcomes [17,18], although this may reduce the scalability of such solutions. The design and deployment of conversational agents, such as chatbots, as virtual coaching solutions to deliver psychoeducational interventions for mental health and well-being have so far been shown to be ideal in maintaining the intuitiveness and naturalness of dialogue-based interaction, while exploiting the benefits of full automation [19,20]. These solutions seem to be particularly interesting to deploy at the time of the COVID-19 pandemic, when restrictions to face-to-face social encounters and interactions make it even more difficult to access human psychological support.

The use of chatbots for digital mental health interventions has attracted interest in the design community. A growing number of studies are reporting their acceptability and feasibility for users [20-22], as well as their effectiveness in reducing perceived stress [16-25] and abnormal eating behavior [26], thereby improving symptoms of anxiety [10,23-25], depression [10,24,25], and insomnia [27]. Like previous work and meta reviews have found on digital mental health interventions for the general adult population and for university students, a main limitation of these studies is their exclusive focus on randomized controlled trials, which prevents a full understanding of the challenges regarding user engagement, uptake, and adoption of these solutions [15-28]. More research is needed to understand the user experience (UX) and engagement with digital mental health interventions, to not only prove their clinical efficacy but also to facilitate their successful implementation in real-world settings [25-29].

The objective of this study was to assess the levels of engagement by university students and effectiveness of their interactions with a psychoeducational intervention delivered by the Atena chatbot over 1 month, in order to improve their coping and resilience skills during the COVID-19 pandemic. The study design was based on a mixed methods approach and encompassed two phases of the ORBIT (Obesity-Related Behavioral Intervention Trials) framework [30] for intervention design (Phase I) and preliminary testing (Phase IIa). In Phase I, the intervention, targets, and components were defined in order to specify their clinically relevant effect on users and to refine the intervention components. In Phase IIa, a proof-of-concept implementation of the digital intervention and chatbot was realized and preliminary testing was done for engagement and effectiveness with a convenience sample of university students. We hypothesized that use of the Atena chatbot over a 1-month period would lead to a reduction in symptoms of stress and anxiety and would prove to be engaging and acceptable to use by students.

To our knowledge, this is the first study to investigate the potential effect of a healthy-coping chatbot intervention during the COVID-19 pandemic, a time when the empowerment gained from stress management skills is much needed by the adult population, particularly university students.

Methods

The Atena Chatbot Design

Atena is a chatbot that delivers psychoeducational content in order to coach users in using coping strategies and improving their mental well-being by means of conversational dialogues with the coach Atena and audiovisual educational materials. It is accessible for free on the Telegram messaging app and is available on mobile or desktop devices. The chatbot was built using JavaScript and was developed by the Digital Health Lab at Fondazione Bruno Kessler (FBK) research center.

The digital mental health intervention delivered by Atena is aimed at improving users' well-being by raising self-awareness about one's thoughts and emotions; it does this by suggesting effective coping strategies that can be adopted when facing typical stressful situations, thus promoting mental well-being and preventing mental distress. The full program consists of eight short sessions, each lasting about 10 minutes, delivered twice a week for 4 weeks. Each session is initiated by the chatbot on a scheduled plan decided by the user during the first session. Users are invited by the chatbot to fill in web-based versions of psychological symptom questionnaires at baseline and postintervention, as well as a user engagement scale at the end of week 2.

The conversations between Atena and the user are informed by evidence-based approaches and intervention strategies from positive psychology and CBT, including psychoeducation on self-awareness and self-efficacy, conflict resolution, assertive communication, and practical exercises on mindfulness delivered

at the end of each session [31,32]. These intervention strategies based on positive psychology, CBT, and mindfulness practices have been recently deployed by fully automated conversational agents targeting anxiety, stress, and depression, with promising results in terms of efficacy and acceptability by users [10,16,24,25]. The aim of the conversations is to trigger and support the user in self-reflecting on personal thoughts and emotions experienced in daily stressful settings, in learning how to best deploy more functional strategies to overcome difficulties, and in better managing stress and anxiety. The intervention program, including the behavioral and clinical targets, as well as the audiovisual content, were originally developed by a team of three clinical psychologists to fit the needs of the general adult population facing stress and anxiety challenges posed by the COVID-19 pandemic (Figure 1). A refinement of the conversations and video materials was then performed by the same psychologists in collaboration with two UX and behavior change experts in the design team, in order to adapt the language, videos, and chatbot interaction to the needs of the students in the target group.

The Atena chatbot always starts the conversation on the scheduled date and time for that session, and the user replies by choosing among a predefined set of answer options. In this way, the conversation flow can be customized to sound more relevant and empathic to the user's answers. Each session starts with a short psychoeducational video cartoon, representing typical challenging situations experienced by young characters and the corresponding strategies adopted to cope with them; they mimic relevant situations experienced by the target users, fostering their identification with those situations and their learning. In the final part of the session, the chatbot invites the user to perform a mindfulness exercise to focus her sense of presence and attention by following a coaching voice provided through an audio track (Figure 2).

Upon enrollment, users were made aware that the Atena chatbot was not intended to replace professional mental health treatment, but that it was a prototyped digital tool designed to support psychoeducational interventions and was going through preliminary testing in this study. In the first session, an introductory video cartoon was presented to the users by the Atena chatbot to explain the main features, applications, and limitations of chatbot technology, in order to facilitate the creation of appropriate expectations toward the digital tool and intervention tested.

Figure 1. Definition of the healthy-coping behavioral intervention and clinical outcome.

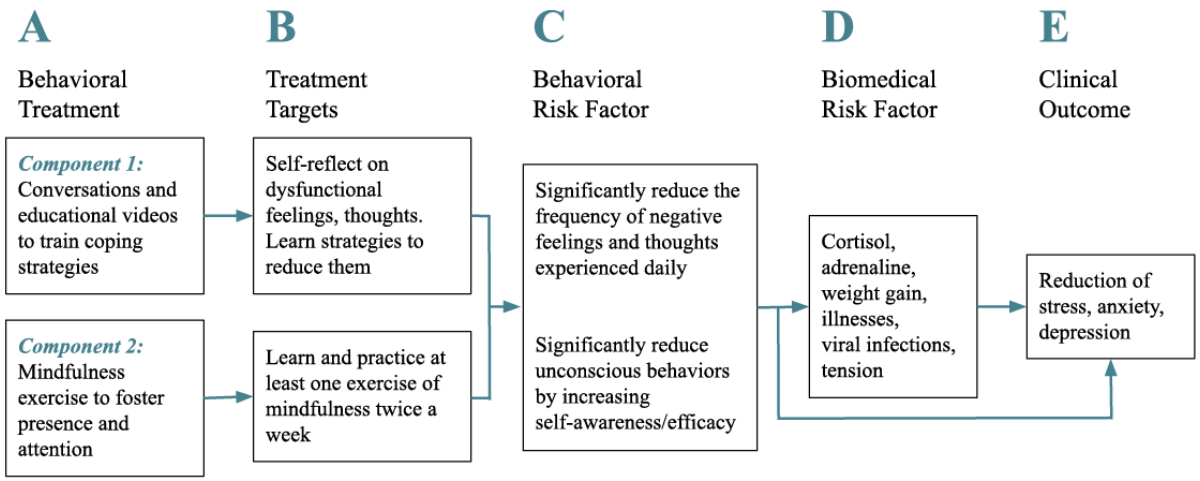
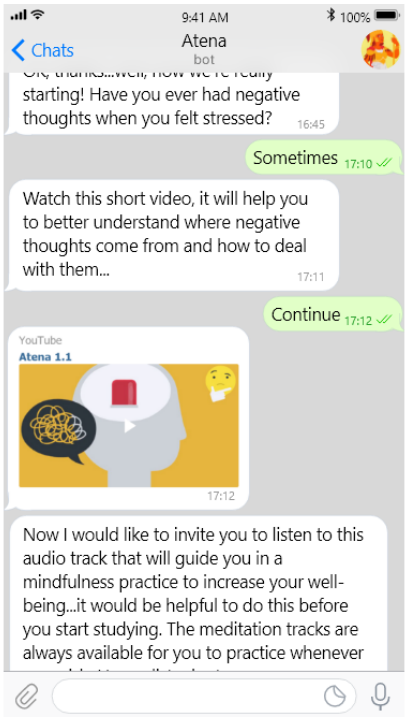


Figure 2. Screenshot from session 1 of the Atena chatbot intervention.



Participants

The Atena chatbot was voluntarily accessed and used by a recruited convenience sample of 71 university students; students ranged in age from 18 to 34 years (mean 20.6, SD 2.4) and the group consisted of 68% females (48/71). Students attended a human-computer interaction course in the first year of a bachelor's degree program at the University of Trento, Italy; they were recruited and invited by SG and RM to access the chatbot via the messaging app Telegram. Participation was on a voluntary basis and the inclusion criteria were as follows: (1) being a university student in their first academic year and (2) owning a smartphone with a Telegram account. Atena was designed to offer coaching conversations, audiovisual materials, and mindfulness meditation in order to improve coping skills and well-being. Students used Atena between mid-October and November 2020, a period affected by the second wave of the COVID-19 pandemic in Italy, with restrictions to citizens' mobility, social distancing, and blended learning recommended at the university. All users were Italian speakers located in the northeast of Italy.

Measures

Perceived Stress Scale

The 10-item Perceived Stress Scale (PSS-10) is a brief self-report instrument containing 10 items rated on a 5-point Likert scale, ranging from 0 (never) to 4 (very often). The PSS-10 measures the perception of stress (ie, the degree to which situations are appraised as stressful) by asking respondents to rate the frequency of their thoughts and feelings related to situations that occurred recently [33] (eg, "In the last month, how often have you felt nervous and stressed?"). Levels of stress are determined based on scores as follows: low (0-13), moderate (14-26), and high (27-40). The PSS-10 is one of the most widely used psychological instruments, reporting good psychometric properties. In this study, the Cronbach α was .84.

Generalized Anxiety Disorder Scale

The 7-item Generalized Anxiety Disorder scale (GAD-7) [34] is a 7-item self-report scale based on a 4-point Likert scale, ranging from 0 (not at all) to 3 (nearly every day). The GAD-7 is used to assess anxiety symptoms over the past 2 weeks (eg, "How often have you been bothered by feeling afraid something awful might happen?"). The total scores are separated into the following four categories to determine the anxiety symptom levels: none (0-4), mild (5-9), moderate (10-14), and severe (≥15) [23,24]. In this study, the Cronbach α was .86.

Five-Facet Mindfulness Questionnaire

The Five-Facet Mindfulness Questionnaire (FFMQ) [35] is a 39-item self-report measure, evaluated on a 5-point Likert scale, ranging from 1 (never or very rarely true) to 4 (very often or always true); the questionnaire assesses the tendency to be mindful in daily life. The five facets are as follows: (1) observing (8 items; eg, "When I'm walking, I deliberately notice the sensations of my body moving"), (2) describing (8 items; eg, "I'm good at finding words to describe my feelings"), (3) acting with awareness (8 items; eg, "When I do things, my mind wanders off and I'm easily distracted"), (4) nonjudging (8 items; eg, "I criticize myself for having irrational or inappropriate emotions"), and (5) nonreactivity (7 items; eg, "I perceive my feelings and emotions without having to react to them"). Individual facet scores range from 8 to 40—except for the nonreactivity facet, which ranges from 7 to 35—with higher scores indicating more mindfulness. The sum of the direct- and reverse-scored items gives a total score that ranges from 39 to 195. In this study, the Cronbach α values were .74 for observing, .92 for describing, .85 for acting with awareness, .91 for nonjudging, .79 for nonreactivity, and .84 for the total score.

User Engagement Scale–Short Form

The User Engagement Scale–Short Form (UES-SF) [36] is a self-report measure comprised of 12 items rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The UES-SF measures the main determinants of adherence and, in particular, it assesses four factors: (1) perceived usability (eg, "Using Atena was frustrating"), (2) aesthetic appeal (eg, "Atena appealed to my senses"), (3) focused attention (eg, "I lost myself in this experience"), and (4) reward (eg, "This experience was rewarding"). The *reward* factor is a summary of three factors from the original User Engagement Scale (UES): *endurability*, a measure of how successful the interaction was and the likelihood of recommending the app to others; *novelty*, a measure of curiosity and interest; and *felt involvement*, a measure of the feeling of being "drawn in" and having fun [36]. A total score is calculated and scores for each of the four subscales are calculated by adding scores for all the items related to their factor and dividing them by the total items.

A back-translation was conducted; thus, all scales were translated into Italian and validated for use within this population. In this study, the Cronbach α values were .66 for perceived usability, .66 for focused attention, .52 for aesthetic appeal, .87 for reward, and .83 for the total scale.

Procedures

After signing and submitting their digital consent form, users were instructed on how to fill in the web-based versions of the PSS-10, the GAD-7, and the FFMQ questionnaires, by using an alphanumeric pseudonymization code decided by them, and how to access the Atena chatbot on the Telegram app. In the first session with the chatbot, users were welcomed and provided with the introductory video on chatbot technology. They were also asked to set a desired day and time for their two weekly sessions with Atena, according to their preferences. The chatbot prompted the user to start a session at the scheduled day and time, but the user was free to pause, continue, or discontinue the session at any time.

At the end of week 2 (session 4), the users were invited by the chatbot to fill in the UES-SF questionnaire to assess their early engagement with the intervention and submit any free-text comments about their UX with the chatbot during the first 2 weeks of interaction.

At the end of week 4 (session 8), the users were invited by the chatbot to again fill in the PSS-10, GAD-7, and FFMQ questionnaires. The chatbot thanked them for their participation in the study and recommended that they continue engaging with the psychoeducational content that had been delivered, in order to improve their coping skills. A total of 4 weeks after the end of the study, participants were also asked to fill in a brief online survey to report on what they most liked and disliked about their experience with Atena, and whether they had continued to practice any of the exercises provided during the intervention with or without the support of Atena in the last 4 weeks. No monetary incentive was provided for participating in this study.

Ethics and Informed Consent

The study was reviewed and approved by the FBK Institutional Ethics Board since it involved a nonclinical population. Participants indicated their consent for their pseudonymized data to be used for research purposes after reading an information sheet. All study data were collected by the Digital Health Lab of FBK. Because of deidentification of all data transmitted between the Atena chatbot and the user, usage data were not linked to specific research participants and were, therefore, reported in an aggregated format.

Statistical Analyses

Statistical analyses were performed using R, version 4.0.0 (The R Foundation) [37], and SPSS Statistics, version 24.0 (IBM Corp) [38]. The Shapiro test was carried out to evaluate the normal distribution of the variables included in this study, such as the GAD-7 and the PSS-10 questionnaire scores (ie, anxiety and stress symptoms, respectively), the total score and the five subscale scores of the FFMQ, and the UES-SF questionnaire score.

The main descriptive analyses (ie, mean, standard deviation, and frequencies) were performed in order to assess the demographic characteristics of the overall sample (ie, age and gender), as well as the GAD-7 and the PSS-10 scores (ie, anxiety and stress symptoms, respectively), the total score and the five subscale scores of the FFMQ, and the UES-SF questionnaire score.

A paired-samples *t* test was conducted to evaluate differences between pre- and postintervention scores concerning the GAD-7 and the PSS-10 (ie, anxiety and stress symptoms, respectively) as well as the total score and the five subscale scores of the FFMQ. A *P* value equal to or less than .05 was considered statistically significant.

An independent-samples *t* test was conducted in order to evaluate differences between pre- and postintervention findings considering two clusters of users' symptoms, extreme versus moderate ranges, as follows: (1) minimal and severe and (2) mild and moderate referring to the GAD-7 and the PSS-10 questionnaire scores. A *P* value equal to or less than .05 was considered statistically significant.

Participants' responses to open-ended questions from the online final survey were analyzed by SR and SC using thematic analysis and were reported as frequencies. Data were analyzed thematically using an inductive, data-driven approach guided by the procedure outlined in Braun and Clarke [39]. Data codes were generated systematically, then collated into themes and applied to the entire data set to generate frequencies.

Results

Participants' Questionnaire Scores by Gender

All the variables included in the analyses were normally distributed. As displayed in Table 1, at baseline, the overall sample (N=71) showed a mean that was close to the moderate range regarding the GAD-7 and the PSS-10 scores. More specifically, regarding the level of anxiety measured with the GAD-7, 35% (25/71) of students were in the *mild* range (ie,

score of 5-9), 31% (22/71) were in the *moderate* range (ie, score of 10-14), and 20% (14/71) were in the *severe* range (ie, score of ≥15). Only 14% (10/71) of students were in the *minimal* range (ie, score of 0-4). Regarding anxiety symptoms for each gender, the analyses show that males experienced mild anxiety symptoms, while females experienced moderate anxiety symptoms.

Regarding stress symptoms evaluated with the PSS-10, 65% (46/71) of the sample were in the *moderate* range (ie, score of 14-26), 7% (5/71) were in the *low* range (ie, score of 0-13), and 28% (20/71) were in the *high* range (ie, score of 27-40). Both males and females displayed moderate stress symptoms.

Regarding the FFMQ scores, participants had an average total mindfulness score of 119.96 (SD 16.99); males scored higher (mean 126.26, SD 19.22) than females (mean 116.93, SD 15.10).

Table 1. Participants' questionnaire scores by gender.

Questionnaire	Questionnaire score, mean (SD)		
	Overall sample (N=71)	Males (n=23)	Females (n=48)
GAD-7[a]	9.92 (4.88)	9.26 (4.66)	10.23 (5.01)
PSS-10[b]	22.46 (6.68)	20.78 (6.65)	23.27 (6.61)
FFMQ[c]			
Observing facet	23.54 (5.81)	23.30 (5.78)	23.15 (5.89)
Describing facet	23.55 (7.16)	25.00 (7.11)	22.85 (7.15)
Act with awareness facet	25.85 (6.10)	26.30 (6.17)	25.63 (6.12)
Nonjudging facet	26.04 (7.95)	28.13 (6.34)	25.04 (8.50)
Nonreacting facet	18.44 (4.40)	20.39 (5.26)	17.50 (3.61)
Total	119.96 (16.99)	126.26 (19.22)	116.93 (15.10)

[a]GAD-7: 7-item Generalized Anxiety Disorder scale. Scores for each item range from 0 (not at all) to 3 (nearly every day), and levels of anxiety are determined based on total scores as follows: none (0-4), mild (5-9), moderate (10-14), and severe (≥15).

[b]PSS-10: 10-item Perceived Stress Scale. Scores for each item range from 0 (never) to 4 (very often), and levels of stress are determined based on total scores as follows: low (0-13), moderate (14-26), and high (27-40).

[c]FFMQ: Five-Facet Mindfulness Questionnaire. Scores for each item within each facet range from 1 (never or very rarely true) to 4 (very often or always true). Individual facet scores range from 8 to 40—except for the nonreacting facet, which ranges from 7 to 35—with higher scores indicating more mindfulness, and total scores range from 39 to 195.

Attrition

A total of 86% (61/71) of participants provided data on the user engagement questionnaire at the end of week 2, which represented an overall attrition rate of 14% (10/71). A total of 58% (41/71) of participants completed the postintervention questionnaire, which represented an overall attrition rate of 42% (30/71).

Dropout was higher among participants in the *minimal* and *mild* anxiety ranges of the GAD-7 questionnaire—50% (5/10) and 44% (11/25), respectively—and was lower in the *moderate* and *severe* anxiety ranges: 40% (9/22) and 35% (5/14), respectively. Moreover, dropout was higher among participants in the *moderate* and *low* stress ranges of the PSS-10 questionnaire—45% (21/46) and 40% (2/5), respectively—and lower in the *high* stress range: 35% (7/20).

User Engagement With the Atena Chatbot

Table 2 shows the results of user engagement with the Atena chatbot as measured by the UES-SF questionnaire at week 2. A total of 61 out of 71 (86%) participants were in agreement with the perceived usability factor, which measured the affective (ie, frustration) and cognitive (ie, effortful) aspects as a result of the interaction. Participants answered in neutral ways regarding (1) the total UES-SF score; (2) the aesthetic appeal factor, which measured the sensory and visual appearance of the interface; and (3) the reward factor, which measured the hedonic aspects of experience, the felt involvement, the overall success of the interaction, and the willingness to engage with the chatbot in the future. Lastly, regarding the focused attention factor, which evaluated the focused concentration, absorption, and temporal dissociation, participants were in disagreement. Since participants were students attending a human-computer interaction course, their expectations regarding the UX and the quality of the user engagement with the chatbot might have been higher compared with students attending other higher education subjects. However, it should also be considered that participants were attending the first semester of their bachelor's degree program, so their expertise in the field of technology design was still quite limited and comparable to that of other students in their age group.

Table 2. User engagement with the chatbot as measured by the UES-SF[a] questionnaire.

UES-SF factor	Score (n=61), mean (SD)
Focused attention factor	2.73 (0.79)
Perceived usability factor	4.28 (0.66)
Aesthetic appeal factor	3.09 (0.65)
Reward factor	3.15 (0.84)
Total	3.15 (0.84)

[a]UES-SF: User Engagement Scale–Short Form. Scores for each item range from 1 (strongly disagree) to 5 (strongly agree). A total score is calculated and scores for each of the four subscales are calculated by adding scores for all the items related to their factor and dividing them by the total items.

Preliminary Efficacy From Completer Analysis

There was a reduction of participants in the *severe* anxiety range of the GAD-7 at postintervention, from 20% (14/71) to 10% (4/41) (Table 3). Also, 7 participants who were above the clinical cutoff score for the GAD-7 of 8 or higher at baseline moved below this cutoff point at postintervention (7/41, 17%).

The independent-samples *t* test between pre- and postintervention between the two clusters of students—cluster 1 students had extreme symptoms and cluster 2 students had moderate symptoms—showed a significant difference (t_{39}=0.94; *P*=.009) among anxiety ranges (ie, GAD-7 scores) in cluster 1, with a decrease of symptoms between preintervention (mean score 12.14, SD 6.88) and postintervention (mean score 10.07, SD 4.58). No other significant difference between pre- and postintervention GAD-7 scores was found.

In line with the GAD-7 results, the PSS-10 scores also showed an increase in the low stress range and a decrease in the high stress range (Table 4). This might indicate a positive effect of the intervention on participants who were in the more extreme stress level ranges compared to those in the intermediate stress ranges. Table 5 shows that the levels of stress symptoms (ie, PSS-10 scores) exhibited significant decreases (t_{39}=2.00; *P*=.05) between pre- and postintervention. Moreover, the mean scores of the subscales *describing* and *nonjudging*, as well as the mean total score of the FFMQ, showed significant increases (*P*<.05) between pre- and postintervention.

Table 3. Classification of anxiety symptoms pre- and postintervention.

GAD-7[a] anxiety level	Students preintervention (N=71), n (%)	Students postintervention (n=41), n (%)
Minimal	10 (14)	5 (12)
Mild	25 (35)	17 (41)
Moderate	22 (31)	15 (37)
Severe	14 (20)	4 (10)

[a]GAD-7: 7-item Generalized Anxiety Disorder scale.

Table 4. Classification of perceived stress symptoms pre- and postintervention.

PSS-10[a] stress level	Students preintervention (N=71), n (%)	Students postintervention (n=41), n (%)
Low	5 (7)	5 (12)
Moderate	46 (65)	27 (66)
High	20 (28)	9 (22)

[a]PSS-10: 10-item Perceived Stress Scale.

Table 5. Paired-samples *t* test between pre- and postintervention (n=41).

Questionnaire	Questionnaire score, mean (SD)		Mean difference (SD)	*t* test (*df*=39)	*P* value[a]
	Preintervention	Postintervention			
GAD-7[b]	10.49 (4.62)	9.29 (0.72)	1.19 (4.14)	1.85	.07
PSS-10[c]	22.49 (6.52)	20.83 (0.97)	1.66 (5.30)	2.00	.05
FFMQ[d]					
Observing facet	23.15 (5.84)	23.37 (6.50)	–0.22 (5.43)	–0.259	.80
Describing facet	23.05 (7.29)	24.98 (6.03)	–1.92 (5.29)	–2.33	.03
Act with awareness facet	26.15 (6.56)	26.12 (6.99)	0.02 (6.27)	0.03	.98
Nonjudging facet	25.85 (7.78)	28.02 (7.46)	–2.17 (6.09)	–2.28	.03
Nonreacting facet	18.41 (4.01)	18.66 (4.85)	–0.24 (4.55)	–0.34	.73
Total	119.49 (16.56)	147.27 (19.67)	–27.78 (16.74)	–10.62	<.001

[a]*P* values were based on two-tailed *t* tests; values were significant at *P*<.05.

[b]GAD-7: 7-item Generalized Anxiety Disorder scale.

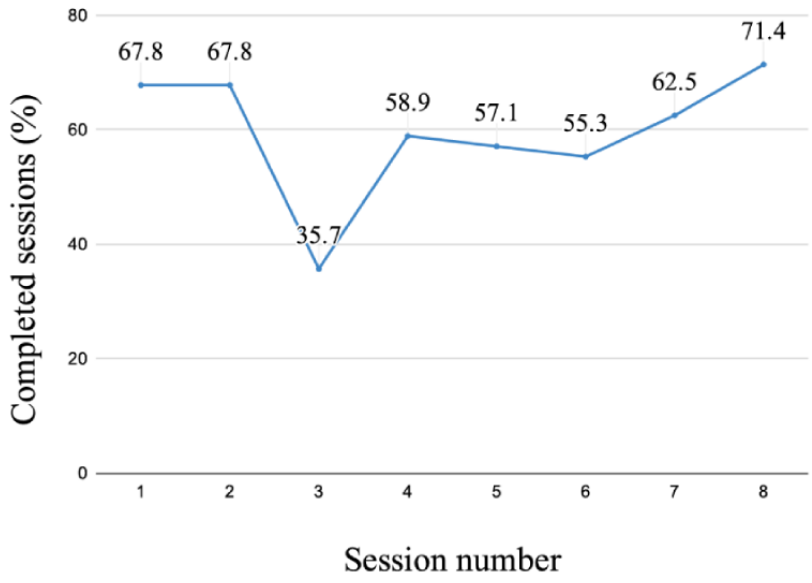
[c]PSS-10: 10-item Perceived Stress Scale.

[d]FFMQ: Five-Facet Mindfulness Questionnaire.

Use of the Chatbot

Participants interacted with the Atena chatbot an average of 78 times (SD 24.8; median 81; range 5-158) over the 4-week period. The average number of uncompleted sessions was 3.1 (SD 2.3) out of 8 overall sessions. Figure 3 shows a graph with the percentage of completed sessions over the 4-week period, showing that engagement and willingness to complete a session was higher during the first and the last weeks of the study.

Figure 3. Overall percentage of completed sessions over the study's 4-week duration.



Qualitative Results

Overview

A total 17 out of 71 (24%) students—65% (11/17) females and 35% males (6/17)—provided qualitative data via the follow-up survey. Three main themes with related subthemes were identified (Table 6): (1) *content*, with the subthemes learning, reflection, multimedia, routine, mindfulness, motivation, originality, and repetitiveness; (2) *UX*, with the subthemes sense of reality, interaction, and flexibility; and (3) *tasks*, with the subthemes notification and availability.

Content

In terms of content, what students liked most was the availability of videos versus only having text-based dialogues in the interaction with the chatbot:

> *It was nice to have the possibility of accessing videos and not just using text, that might be boring.*

However, one participant suggested that videos should be improved from a graphic design point of view:

> *What I liked less were the videos. They were interesting, but I think they should be improved in terms of graphics and visuals in order to be more engaging for users.*

Some students appreciated the opportunity to learn new things—"I appreciated tips provided during the course, which were very interesting and also useful for learning new skills"—and to approach the mindfulness practice:

> *I find the bot an excellent first opportunity to start mindfulness practices, especially for those who - like me - can never find time to stop and breathe and have never tried anything like this before...*

Some students also appreciated the originality of the exercises:

> *The originality of the exercises, in my opinion, is very important to foster change in people and help them.*

One student suggested that the chatbot questions should be more varied in format so as not to be too repetitive:

> *In my opinion, questions should change over time: "how are you" should be asked in a more nuanced way, otherwise it sounds repetitive. I suggest, if possible, to vary the dialogues with the user, especially the welcoming messages.*

User Experience

Regarding the UX, one student reported that what they liked most was the feeling of real-life interaction with the chatbot:

> *First of all, I appreciated the continuity of the short course with the chatbot, the interactions were well thought out and articulated, it felt like chatting with a real person.*

Some students, however, reported some criticism regarding the user-chatbot interaction and gave some suggestions for future improvements:

> *The interaction with Atena should be more personalized based on the user's answers, sometimes the answering options did not take into account the different nuances of mood.*

Tasks

In the task theme, participants' remarks mainly concerned the chatbot notifications and lack of reminders to resume a session when the user was interrupted by some other task:

> *I often didn't have time to watch the videos at the scheduled time, then I forgot to resume them because there was no reminder.*

The possibility of having the materials provided by Atena available in the chat was much appreciated:

> *I can watch videos again whenever I like.*

A total of 8 out of the 17 (47%) students who provided qualitative data (ie, 4 females and 4 males) reported that they accessed the Atena videos and materials again after the 1-month study duration.

Table 6. Themes, subthemes, and quotes by participants.

Main theme and subthemes	Participant quotes
Content	
Learning	"I enjoyed learning about new forms of interaction and their applications." (Participant #3)
	"I appreciated tips given during the course, since they were very interesting and also useful for learning new skills." (Participant #10)
Reflection	"Reported topics were all interesting and quick to address. I spent very little time with the chatbot but I think I got some food for thought." (Participant #1)
Multimedia	"I liked the guided meditation videos." (Participant #5)
	"I really liked the relaxation videos." (Participant #7)
	"It was nice to have the possibility to access videos and not just texts, that can be boring." (Participant #8)
	"I enjoyed the guided meditation videos." (Participant #13)
	"What I liked most were the mindfulness and motivational videos." (Participant #16)
	"Videos were long and I often tended to quit them before the end." (Participant #3)
	"Videos often recommended walking or covering distances, which made it difficult to perform the task for people like me who live in a small indoor place (I often did not have the possibility or desire to move outdoors)." (Participant #9)
	"What I liked less was the videos. They were interesting, but I think they should be improved in terms of graphics and visuals in order to be engaging for users." (Participant #10)
Routine	"Unfortunately receiving videos for me had become a pleasant habit, so when it went over it looked like a sudden interruption to me." (Participant #16)
Mindfulness	"Breathing tips, calm tone of voice put me in a good mood, dialogues were motivating and relaxing." (Participant #9)
	"I find the bot an excellent first opportunity to practice mindfulness, especially for those who - like me - can never find the time to stop and breathe and have never tried anything like this before." (Participant #13)
Motivation	"Low incentive." (Participant #2)
	"It didn't appeal to me so much, so I struggled to be constant." (Participant #15)
Originality	"The originality of the exercises, in my opinion, it is very important to foster a change in people to help them." (Participant #6)
Repetitiveness	"In my opinion, questions should vary over time: 'how are you' should be asked in a more nuanced way, otherwise it sounds repetitive. I suggest, if possible, to vary the dialogues with the user, especially the welcoming messages." (Participant #16)
User experience	
Sense of reality	"First of all, I appreciated the continuity of the short course with the chatbot, the interactions were well thought out and articulated, it felt like chatting with a real person." (Participant #1)
Interaction	"I would have liked to have more dialogues with the chatbot and less external interaction (YouTube video)." (Participant #1)
	"I would prefer to interact and chat whenever I wish, and not only on fixed days." (Participant #8)
	"The interaction with Atena should be more personalized based on the user's answers, sometimes the answering options did not take into account the different nuances of mood." (Participant #9)
	"Interactions could only take place on the days agreed upon at the outset." (Participant #13)
Flexibility	"I really liked being able to have flexibility in the scheduling." (Participant #2)
Tasks	
Notification	"Notifications reminding people to take the test." (Participant #4)
	"I feel satisfied. My only remark is about the notifications in the dialogue flow, which tended to be overlooked." (Participant #4)
	"Not having time when the message arrived and then forgetting to do the exercises." (Participant #5)
	"Atena sent me notifications when it was not suitable for me (despite my choosing of scheduling options) and then I forgot to do the activity." (Participant #7)
	"It would be nice to specify at the very beginning when an activity requires places larger than a room or also to be outdoors." (Participant #9)
	"I often didn't have time to watch videos at the scheduled time, then I forgot to resume them because there was no reminder." (Participant #15)

Main theme and subthemes	Participant quotes
Availability	"I can watch videos again whenever I like." (Participant #11)
	"The scheduling options, in my opinion, should be more restrictive, or maybe deadlines could be set for some tasks." (Participant #6)
	"The videos made it difficult for me to find enough time to watch them." (Participant #12)

Discussion

Principal Findings

Results from this preliminary evaluation of the Atena chatbot intervention indicate that healthy-coping psychoeducation can be effectively deployed to university students and can have positive effects, especially on those who are more in need of psychological support to cope with stress and anxiety symptoms. Our results are in line with recent studies targeting the same population, showing that online stress management interventions are more effective for students with higher levels of stress, anxiety, and depression [40,41]. Higher engagement and lower attrition rates were also observed in those students in our sample who had more severe levels of anxiety at baseline. This is a very promising result for implementing future anxiety prevention and management solutions to be delivered during the COVID-19 pandemic and beyond. Results also showed a significant improvement in the capacity of participants to describe and accept their emotions, which can be an effect of the mindfulness practice and self-reflection elicited by the conversations with the chatbot. Training in these kinds of skills may be particularly needed by the university student population and could have positive effects on students' mental well-being.

The baseline levels of stress and anxiety among our participants were significantly higher than those found in the same population of university students by previous research [42-44], as well as in the general population [45]. Previous studies analyzing the mental health of university students found lower levels of anxiety and stress symptoms, evaluated through the GAD-7 and the PSS-10 questionnaires, compared to those of our sample of university students [42-44]. This is not surprising, since the COVID-19 pandemic had a worsening effect on the general population and on university students in particular: a research study conducted on Italian university students to identify psychological consequences of the living conditions during the COVID-19 lockdown reported high levels of anxiety and stress, concentration disorders, psychosomatization, and, in several cases, reactivation of trauma and worsened sleep quality [46]. Moreover, from the time of our baseline assessment until the end of our study, the COVID-19 pandemic in Italy reached higher levels of infection, which required more severe restrictions to be introduced in schools and universities (ie, full online teaching) and in citizens' everyday lives. It is likely that students experiencing more severe symptoms of anxiety felt more motivated to engage with the Atena chatbot and found it a more convenient solution to access psychoeducational support, by avoiding, at the same time, stigma and possible difficulties in accessing mental health services.

Overall, our quantitative and qualitative findings are aligned with recent chatbot evaluation studies [10,23-25] and will inform our design decisions for future developments. Results regarding attrition rates during the study, user engagement, as well as our qualitative findings suggest that our proof-of-concept intervention needs to be further refined to fully meet the requirements and preferences of the target users before being ready for randomized controlled trial evaluations. Our analysis indicates that the engagement and attractiveness of a chatbot-based mental health intervention for university students might wear off or reduce significantly after 2 weeks of interaction, requiring deeper levels of engagement through conversation and rewarding feedback from the chatbot in order to maintain users' interest and commitment during the intervention. This might be particularly useful for supporting adherence of less motivated users, such as the ones with mild to moderate levels of stress and anxiety. Our study also helps shed light on what might be the ideal length, frequency, or intensity of digital mental health interventions for nonclinical populations. The duration and intensity of our intervention was sufficient to provide psychoeducational support to students without interfering too much with their daily life commitments, and it was also effective in triggering more self-reflection and mindfulness practice in the follow-up period. This might be interpreted as a signal of user empowerment and desired behavior change, although more research is needed to confirm this interpretation.

Limitations

This study presents some limitations that affect the generalizability of the findings. It reports data from the preliminary evaluation of a proof-of-concept chatbot intervention targeting a homogeneous population of university students without a control group. Although the findings on user engagement and preliminary effectiveness of the intervention are promising and aligned with previous research, further testing by means of controlled trials should be conducted to confirm any conclusion about efficacy and to verify its maintenance at follow-up. However, the evidence presented from students' responses and feedback to the intervention confirm the importance of deploying user-centered methodology in the iterative design and refinement of these interventions before investing additional resources in conducting more rigorous efficacy testing.

Another limitation is related to our method of collecting objective data on users' engagement and interaction with the chatbot intervention during the study. Since our log data were deidentified, it was more difficult to assess any difference among users in how deeply they focused attention and self-reflected upon the psychoeducational videos' contents and chatbot suggestions during each session. Although we could derive some information on users' satisfaction with these contents from participants' qualitative comments, a more complete and objective monitoring of users' behavioral interactions with the

intervention's components would be preferable to deploy in future studies.

Finally, the study was conducted during the second wave of the COVID-19 pandemic in Italy, characterized by the introduction of increasingly more rigid restrictions to social behavior and educational practices that might have strongly impacted the mental well-being of our participants and reduced the positive effect of our intervention. However, the challenging contextual setting in which our intervention was deployed can also be considered a point of strength of the contribution provided by this study, offering interesting insights for the future wider deployment of digital mental health interventions in challenging conditions.

Conclusions

This study further extends previous research on the use of chatbot-based interventions for healthy coping with stress, confirming their effectiveness in supporting university students experiencing higher levels of distress. Although the generalizability of the reported findings should be viewed with caution, since no control group was involved and the intervention was deployed during the COVID-19 pandemic, these preliminary findings are interesting for inspiring the future design of digital mental health interventions for university students and public health.

Acknowledgments

Special thanks go to the students who participated in the study. Thanks also go to Letizia Esposti and Griselda Mahmutaj for their support and comments during the study.

Authors' Contributions

SG, SR, and GB contributed substantially to the conception and design of the study, the acquisition of data, and writing of the paper. SF and SC provided critical review and significant contributions to the manuscript. RM and MM contributed to the development of the Atena chatbot and to the acquisition of the log data. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

All the authors are employees of Digital Health Lab, Fondazione Bruno Kessler that developed the Atena chatbot.

References

1. Pinder-Amaker S, Bell C. A bioecological systems approach for navigating the college mental health crisis. Harv Rev Psychiatry 2012;20(4):174-188. [doi: 10.3109/10673229.2012.712842] [Medline: 22894727]
2. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by US college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
3. Le HT, Lai AJX, Sun J, Hoang MT, Vu LG, Pham HQ, et al. Anxiety and depression among people under the nationwide partial lockdown in Vietnam. Front Public Health 2020 Oct 29;8:589359 [FREE Full text] [doi: 10.3389/fpubh.2020.589359] [Medline: 33194995]
4. Tran BX, Nguyen HT, Le HT, Latkin CA, Pham HQ, Vu LG, et al. Impact of COVID-19 on economic well-being and quality of life of the Vietnamese during the national social distancing. Front Psychol 2020;11:565153 [FREE Full text] [doi: 10.3389/fpsyg.2020.565153] [Medline: 33041928]
5. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: A review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
6. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
7. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
8. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: Results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
9. Fukuda CC, Penso MA, Amparo DMD, Almeida BCD, Morais CDA. Mental health of young Brazilians: Barriers to professional help-seeking. Estud Psicol (Campinas) 2016 Jun;33(2):355-365. [doi: 10.1590/1982-02752016000200017]
10. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

11. Menezes P, Quayle J, Garcia Claro H, da Silva S, Brandt LR, Diez-Canseco F, et al. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: A pilot study in Brazil and Peru. JMIR Ment Health 2019 Apr 26;6(4):e11698 [FREE Full text] [doi: 10.2196/11698] [Medline: 31025949]
12. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: A survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014 Dec;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
13. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: A systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
14. Zhang MW, Ho RC. Moodle: The cost effective solution for internet cognitive behavioral therapy (I-CBT) interventions. Technol Health Care 2017 Feb 21;25(1):163-165. [doi: 10.3233/thc-161261]
15. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
16. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
17. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-869; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
18. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — A systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
19. Suganuma S, Sakamoto D, Shimoyama H. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: Feasibility and acceptability pilot trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]
20. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
21. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: Mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
22. Gabrielli S, Rizzi S, Carbone S, Donisi V. A chatbot-based coaching intervention for adolescents to promote life skills: Pilot study. JMIR Hum Factors 2020 Feb 14;7(1):e16762 [FREE Full text] [doi: 10.2196/16762] [Medline: 32130128]
23. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
24. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
25. Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M. Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. Front Digit Health 2020 Nov 30;2:1-7. [doi: 10.3389/fdgth.2020.576361]
26. Low TL, Ho R, Ho C, Tam W. The efficacy of virtual reality in the treatment of binge-purging eating disorders: A meta-analysis. Eur Eat Disord Rev 2021 Jan 12;29(1):52-59. [doi: 10.1002/erv.2804] [Medline: 33180339]
27. Soh HL, Ho RC, Ho CS, Tam WW. Efficacy of digital cognitive behavioural therapy for insomnia: A meta-analysis of randomised controlled trials. Sleep Med 2020 Nov;75:315-325. [doi: 10.1016/j.sleep.2020.08.020] [Medline: 32950013]
28. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: Systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
29. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: Systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
30. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
31. Seligman MEP. Building human strength: Psychology's forgotten mission. APA Monitor. 1998. URL: https://nonopp.com/ar/Psicologia/00/pres.htm [accessed 2021-05-25]
32. Beck JS. Cognitive Behavior Therapy: Basics and Beyond. 2nd edition. New York, NY: The Guilford Press; 2011.
33. Cohen S. Perceived Stress Scale. Mind Garden. 1994. URL: http://mindgarden.com/documents/PerceivedStressScale.pdf [accessed 2020-05-17]
34. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

35. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
36. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
37. R Core Team. R 4.0.0 for Windows. The Comprehensive R Archive Network. Vienna, Austria: The R Foundation; 2020. URL: https://cran.r-project.org/bin/windows/base/old/4.0.0/ [accessed 2021-05-25]
38. IBM SPSS Statistics for Windows, Version 24.0. IBM Corp. Armonk, NY: IBM Corp; 2016. URL: https://www.ibm.com/docs/en/spss-statistics/24.0.0 [accessed 2021-05-25]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Coudray C, Palmer R, Frazier P. Moderators of the efficacy of a web-based stress management intervention for college students. J Couns Psychol 2019 Nov;66(6):747-754. [doi: 10.1037/cou0000340] [Medline: 30998054]
41. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: Meta-analysis of individual patient data. BMJ 2013 Feb 26;346(feb26 2):f540-f540. [doi: 10.1136/bmj.f540] [Medline: 23444423]
42. Mirón J, Goldberg X, López-Solà C, Nadal R, Armario A, Andero R, et al. Perceived stress, anxiety and depression among undergraduate students: An online survey study. J Depress Anxiety 2019;8(1):330 [FREE Full text] [doi: 10.4172/2167-1044.1000330]
43. Alghadir A, Manzar MD, Anwer S, Albougami A, Salahuddin M. Psychometric properties of the Generalized Anxiety Disorder Scale among Saudi university male students. Neuropsychiatr Dis Treat 2020 Jun;16:1427-1432 [FREE Full text] [doi: 10.2147/ndt.s246526]
44. Anwer S, Manzar MD, Alghadir AH, Salahuddin M, Abdul Hameed U. Psychometric analysis of the Perceived Stress Scale among healthy university students. Neuropsychiatr Dis Treat 2020 Oct;16:2389-2396 [FREE Full text] [doi: 10.2147/ndt.s268582]
45. Wang C, Tee M, Roy AE, Fardin MA, Srichokchatchawan W, Habib HA, et al. The impact of COVID-19 pandemic on physical and mental health of Asians: A study of seven middle-income countries in Asia. PLoS One 2021 Feb 11;16(2):e0246824 [FREE Full text] [doi: 10.1371/journal.pone.0246824] [Medline: 33571297]
46. Savarese G, Curcio L, D'Elia D, Fasano O, Pecoraro N. Online university counselling services and psychological problems among Italian students in lockdown due to Covid-19. Healthcare (Basel) 2020 Oct 29;8(4):440 [FREE Full text] [doi: 10.3390/healthcare8040440] [Medline: 33137963]

Abbreviations

CBT: cognitive behavioral therapy
FBK: Fondazione Bruno Kessler
FFMQ: Five-Facet Mindfulness Questionnaire
GAD-7: 7-item Generalized Anxiety Disorder scale
ORBIT: Obesity-Related Behavioral Intervention Trials
PSS-10: 10-item Perceived Stress Scale
UES: User Engagement Scale
UES-SF: User Engagement Scale–Short Form
UX: user experience

Edited by L Buis; submitted 15.02.21; peer-reviewed by R Ho, A Coe; comments to author 14.03.21; revised version received 01.04.21; accepted 20.04.21; published 28.05.21.

Please cite as:
Gabrielli S, Rizzi S, Bassi G, Carbone S, Maimone R, Marchesoni M, Forti S
Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study
JMIR Mhealth Uhealth 2021;9(5):e27965
URL: https://mhealth.jmir.org/2021/5/e27965
doi:10.2196/27965
PMID:33950849

Original Paper

Performance of a Mobile Single-Lead Electrocardiogram Technology for Atrial Fibrillation Screening in a Semirural African Population: Insights From "The Heart of Ethiopia: Focus on Atrial Fibrillation" (TEFF-AF) Study

Bradley M Pitman[1], BSc; Sok-Hui Chew[2], BMBS; Christopher X Wong[1], MSc, MBBS, PhD; Amenah Jaghoori[1], PhD; Shinsuke Iwai[1], MD, PhD; Gijo Thomas[1], PhD; Andrew Chew[2], BMBS; Prashanthan Sanders[1], MBBS, PhD; Dennis H Lau[1], MBBS, PhD

[1]Centre for Heart Rhythm Disorders, The University of Adelaide, Adelaide, Australia
[2]Flinders University, Adelaide, Australia

Corresponding Author:
Dennis H Lau, MBBS, PhD
Centre for Heart Rhythm Disorders
The University of Adelaide
1 Port Rd
Adelaide, 5000
Australia
Phone: 61 8313 9000
Fax: 61 8362 2273
Email: dennis.h.lau@adelaide.edu.au

Abstract

Background: Atrial fibrillation (AF) screening using mobile single-lead electrocardiogram (ECG) devices has demonstrated variable sensitivity and specificity. However, limited data exists on the use of such devices in low-resource countries.

Objective: The goal of the research was to evaluate the utility of the KardiaMobile device's (AliveCor Inc) automated algorithm for AF screening in a semirural Ethiopian population.

Methods: Analysis was performed on 30-second single-lead ECG tracings obtained using the KardiaMobile device from 1500 TEFF-AF (The Heart of Ethiopia: Focus on Atrial Fibrillation) study participants. We evaluated the performance of the KardiaMobile automated algorithm against cardiologists' interpretations of 30-second single-lead ECG for AF screening.

Results: A total of 1709 single-lead ECG tracings (including repeat tracing on 209 occasions) were analyzed from 1500 Ethiopians (63.53% [953/1500] male, mean age 35 [SD 13] years) who presented for AF screening. Initial successful rhythm decision (normal or possible AF) with one single-lead ECG tracing was lower with the KardiaMobile automated algorithm versus manual verification by cardiologists (1176/1500, 78.40%, vs 1455/1500, 97.00%; P<.001). Repeat single-lead ECG tracings in 209 individuals improved overall rhythm decision, but the KardiaMobile automated algorithm remained inferior (1301/1500, 86.73%, vs 1479/1500, 98.60%; P<.001). The key reasons underlying unsuccessful KardiaMobile automated rhythm determination include poor quality/noisy tracings (214/408, 52.45%), frequent ectopy (22/408, 5.39%), and tachycardia (>100 bpm; 167/408, 40.93%). The sensitivity and specificity of rhythm decision using KardiaMobile automated algorithm were 80.27% (1168/1455) and 82.22% (37/45), respectively.

Conclusions: The performance of the KardiaMobile automated algorithm was suboptimal when used for AF screening. However, the KardiaMobile single-lead ECG device remains an excellent AF screening tool with appropriate clinician input and repeat tracing.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619001107112; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=378057&isReview=true

(JMIR Mhealth Uhealth 2021;9(5):e24470) doi:10.2196/24470

KEYWORDS

atrial fibrillation; screening; sub-Saharan Africa; single-lead ECG

Introduction

Consumer use of wearable technology capable of ambulatory assessment of heart rate and rhythm has significantly increased in recent years [1]. Large-scale population screening studies have demonstrated the capability of wearable devices to detect pulse irregularity using photoplethysmography-based technology, with a high positive predictive value of diagnosing atrial fibrillation (AF) [2,3]. However, the adoption of these smart wearable devices is much lower in low-resource countries due to affordability and low internet penetration rate. Despite AF being recognized as a growing global epidemic, the 2010 Global Burden of Disease study has highlighted low availability of data on AF from several regions including sub-Saharan Africa and the need for better estimates through targeted population surveillance studies [4]. Alternative active screening strategies for AF using pulse palpation and electrocardiogram (ECG) are therefore more applicable in these low-resource countries [1,5].

AF screening using single-lead ECG devices has been reported in hospital, primary care, and community settings with variable sensitivity and specificity [6]. However, limited data exist on the use of such devices for AF screening in low-resource countries [7]. One such device is the KardiaMobile ECG monitor (AliveCor Inc), which is approved by the US Food and Drug Administration (FDA) for automatic classification of 30-second single-lead ECG tracing as normal or possible AF. However, the device also returns other results of too short, tachycardia, bradycardia, unreadable, or unclassified. Notably, screening studies using the KardiaMobile device, including the Heart Rhythm Society/American College of Physicians AF Screening and Education Initiative, have encountered between 5% and 28% of unclassified ECG recordings [8-12]. The high frequency of unclassified tracings may limit the effective utility of this device for AF screening. Here, we sought to determine the real-world feasibility and utility of the KardiaMobile single-lead ECG device for AF screening in a semirural African population. Specifically, this analysis evaluates the device's accuracy for AF detection, factors underlying unclassified ECG tracings, and factors that may influence its screening performance from the first 1500 subjects recruited in the ongoing TEFF-AF (The Heart of Ethiopia: Focus on Atrial Fibrillation study).

Methods

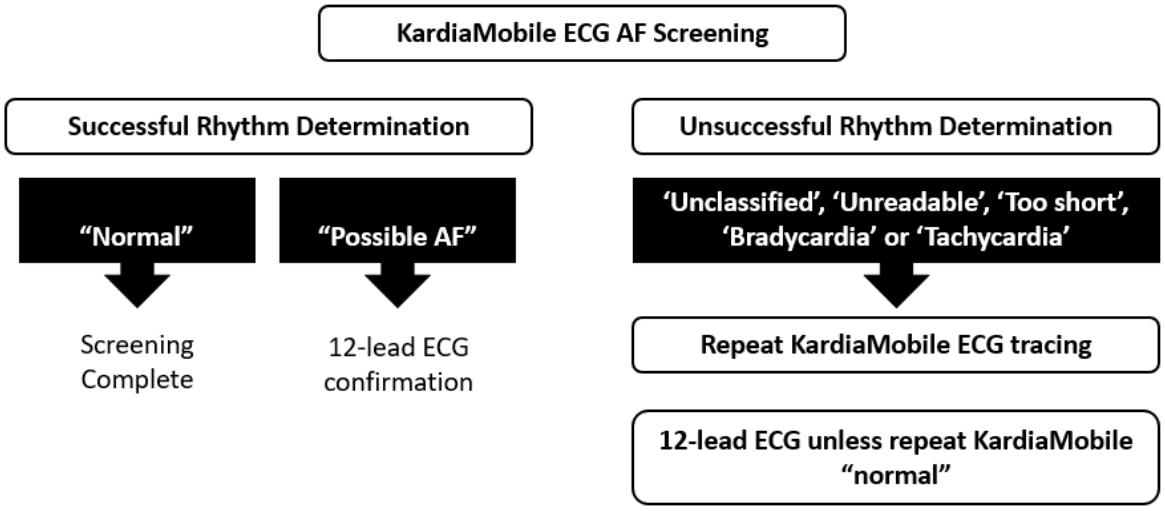
TEFF-AF Study

The TEFF-AF study (registered with the Australian New Zealand Clinical Trials Registry [ACTRN12619001107112]) is an AF screening study conducted at the Soddo Christian Hospital (SCH). The SCH is located in the semirural town of Soddo in south-central Ethiopia, with a population of around 200,000 individuals. AF screening was undertaken by a team of 5 nursing and research support staff from the SCH following specialized training on the use of the KardiaMobile device, iPhone app (version 5.7.4, KardiaAI: 1.1.7), and online Research Electronic Data Capture database. The training included an initial tutoring session followed by subsequent hands-on practice in acquiring a best-quality single-lead ECG tracing with the KardiaMobile device. AF screening commenced at the SCH in August 2019 with inclusion criteria being any ambulant adult aged 18 years and above and able to provide informed consent. Signage in Amharic language was erected to advertise screening to aid recruitment (Figure 1).

Figure 1. Atrial fibrillation screening advertising (left) and study information (center, in Amharic language) and single-lead electrocardiogram recording (right).

All participants provided informed consent, and the study is approved by the SCH research ethics board. Baseline demographic and clinical parameters were obtained to characterize the cardiovascular risk profile of participating individuals. Measurements of height, weight, and blood pressure (Intellisense T5 automatic monitor, Omron Corporation) were obtained before single-lead ECG acquisition using the KardiaMobile device. As per the study protocol (Figure 2), the outcome of the automated algorithm assessment of rhythm dictated the need for repeat KardiaMobile tracing and/or a 12-lead ECG. Participants with clinical abnormality detected were referred for follow-up by the SCH physician.

Figure 2. Atrial fibrillation (AF) screening protocol. ECG: electrocardiogram.



KardiaMobile ECG Screening

The KardiaMobile mobile single-lead ECG device records a bipolar lead I ECG tracing when 2 or 3 fingers from each hand of the user are placed in contact with the 2 electrodes (Figure 1). Participants were instructed to relax arms and hands to reduce noise and artefacts. The KardiaMobile device transmits a frequency modulated ultrasound signal that is detected by the smartphone (iPhone, Apple Inc) with installed Kardia app. A 30-second single-lead ECG recording can be viewed in real time on the smartphone app and saved as a PDF file. The noise-filtered trace and computer-averaged complex on the KardiaMobile app is then subjected to an automated algorithm for arrhythmia diagnosis using the 2 criteria of p-wave absence and R-R interval irregularity [13].

ECG Adjudication Analysis

The KardiaMobile ECG tracings obtained for the first consecutive 1500 participants in the TEFF-AF study were included in this analysis. Each single-lead ECG tracing has a rhythm determination by the KardiaMobile automated algorithm of normal, possible AF, bradycardia, tachycardia, unclassified, unreadable, or too short. Single-lead ECG traces were downloaded and analyzed independently by two cardiologists. The cardiologists also assessed diagnostic limitations for each tracing categorized as artefact, ectopy, bradycardia, tachycardia, or insufficient sample duration.

Data Availability

The dataset with deidentified information generated and analyzed during this study is available from the corresponding author on reasonable request.

Statistical Analysis

Summary statistics were presented by frequency and percentage or mean and standard deviation as appropriate. Categorical data were analyzed using the chi-square test. Sensitivity and specificity for the ability of the KardiaMobile to produce a rhythm decision against the cardiologist ECG interpretation was calculated. Linear regression analysis was performed to assess the factors contributing to screening performance of the KardiaMobile automated algorithm. All statistics were performed in SPSS Statistics version 26 (IBM Corp), and statistical significance set at $P<.05$.

Results

Participants

A total of 1709 single-lead ECG tracings (including repeat tracing on 209 occasions) were analyzed from a cohort of 1500 participants who presented for AF screening. The baseline clinical parameters of the participants are shown in Table 1. The mean age was 35 (SD 13) years and 63.53% (953/1500) were male. Of these participants, 95.93% (1439/1500) were from the regional state of Southern Nations, Nationalities, and Peoples' Region where the SCH is located, and 87.07% (1306/1500) had secondary level education or above. The self-reported clinical history of the participants is shown in Table 1, with hypertension (104/1500, 6.93%) as the most prevalent comorbidity.

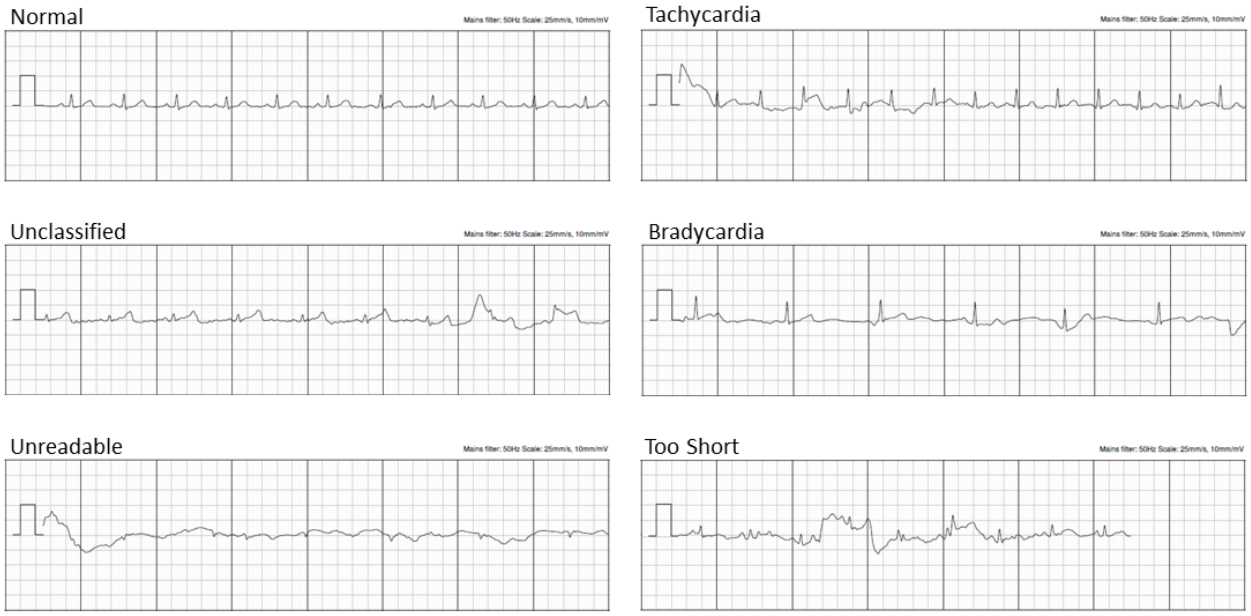
Table 1. Baseline clinical characteristics (n=1500).

Demographic and clinical information	Values
Age in years, mean (SD)	35 (13)
Gender, male, n (%)	960 (64.00)
Home region, n (%)	
Southern Nations, Nationalities, and Peoples' Region	1439 (95.93)
Omoria	30 (2.00)
Amhara	11 (0.73)
Other regions (including Somalia, B-Gumuz, Addis Ababa, Harar)	19 (1.27)
Religion, n (%)	
Orthodox	416 (27.73)
Protestant	988 (65.87)
Muslim	70 (4.67)
Other or no religion	22 (1.47)
Education, n (%)	
Illiterate	55 (3.67)
Primary level school	137 (9.13)
Secondary level school	599 (39.93)
Certificate, diploma, or higher	707 (47.13)
Occupation, n (%)	
Unemployed	175 (11.67)
Employed	682 (45.47)
Self-employed	344 (22.93)
Others including student and retired	297 (19.80)
Clinical, mean (SD)	
Height (cm)	167.7 (8.6)
Weight (kg)	67.1 (13.3)
Systolic blood pressure (mm Hg)	124.0 (17.7)
Diastolic blood pressure (mm Hg)	76.5 (11.7)
Clinical, n (%)	
Hypertension	104 (6.93)
Diabetes mellitus	34 (2.27)
Congestive cardiac failure	20 (1.33)
Stroke	3 (0.20)
Coronary artery disease	2 (0.13)
Peripheral artery disease	0 (0.00)
Chronic lung disease	16 (1.07)
Chronic renal disease	5 (0.33)
Valvular heart disease	11 (0.73)
Obstructive sleep apnea	2 (0.13)
Thyroid disease	21 (1.40)
Smoker	5 (0.33)
Khat/alcohol use	14 (0.93)
Infectious disease	288 (19.20)

Performance of the KardiaMobile Automated Algorithm

Of the initial single-lead ECG tracings from 1500 participants, the KardiaMobile algorithm was unable to provide a rhythm decision in 21.60% (324/1500) due to unclassified (130/1500, 8.67%), tachycardia (128/1500, 8.53%), unreadable (62/1500, 4.13%), too short (3/1500, 0.20%), and bradycardia (1/1500, 0.07%). Representative examples of these tracings are shown in Figure 3. A repeat KardiaMobile tracing was obtained in 64.50% (209/324) of the participants who did not have an initial rhythm decision. Of those participants without repeat KardiaMobile tracings, 83.48% (96/115) had an initial result of tachycardia, which the screening team deemed as sinus tachycardia (>100 bpm) and interpreted as normal rhythm not requiring a repeat tracing, 10.43% (12/115) proceeded directly to a 12-lead ECG, and 6.09% (7/115) declined repeat KardiaMobile tracing or 12-lead ECG due to time constraint. On the repeat KardiaMobile attempt, the KardiaMobile algorithm again failed to achieve a rhythm decision in 40.19% (84/209). Adjudications by cardiologists showed that the reasons underlying unsuccessful automated KardiaMobile rhythm determination (n=408 traces; 324 from first attempt and 84 from repeat attempt) were poor quality/noisy tracings (214/408, 52.45%), tachycardia (>100 bpm; 167/408, 40.93%), frequent ectopy (22/408, 5.39%), inadequate recording duration (3/408, 0.74%), and bradycardia (<50 bpm; 2/408, 0.49%).

Figure 3. Examples of KardiaMobile single-lead electrocardiogram tracings.



KardiaMobile Automated Algorithm Versus Cardiologist Adjudication

The KardiaMobile automated algorithm successfully obtained a rhythm decision on the first attempt for 78.40% (1176/1500) of participants, which was considerably lower than manual assessment by cardiologists (1455/1500, 97.00%; *P*<.001; Figure 4). The sensitivity and specificity of a rhythm decision by the KardiaMobile automated algorithm from the initial single-lead ECG of each participant, when compared with manual assessment by cardiologists, was 80.3% (95% CI 78.1% to 82.3%) and 82.2% (95% CI 68.0% to 92.0%), respectively (Table 2). The KardiaMobile automated algorithm's success in rhythm decision improved to 86.73% (1301/1500) with the inclusion of repeat KardiaMobile tracings achieving a rhythm decision for an additional 125 participants, although it remained lower than manual assessment by cardiologists (1479/1500, 98.60%; *P*<.001; Figure 4). In total, 96.96% (1657/1709) of the single-lead ECG tracings were of adequate quality for diagnostic purposes according to cardiologist adjudication. Notably, all the KardiaMobile algorithm-determined normal single-lead ECG tracings were confirmed to be normal sinus rhythm according to cardiologist adjudication. However, 3 traces that failed to achieve a rhythm decision by KardiaMobile (2 unreadable and 1 unclassified) were deemed AF by cardiologist adjudication. The sensitivity and specificity of AF detection by the KardiaMobile automated algorithm from 1709 single-lead ECG tracings, when compared with manual assessment by cardiologists, was 75.0% (95% CI 42.8% to 94.5%) and 96.4% (95% CI 95.4% to 97.2%), respectively (Table 2).

Figure 4. Comparison of KardiaMobile algorithm versus manual assessment by cardiologists. ECG: electorcardiogram.

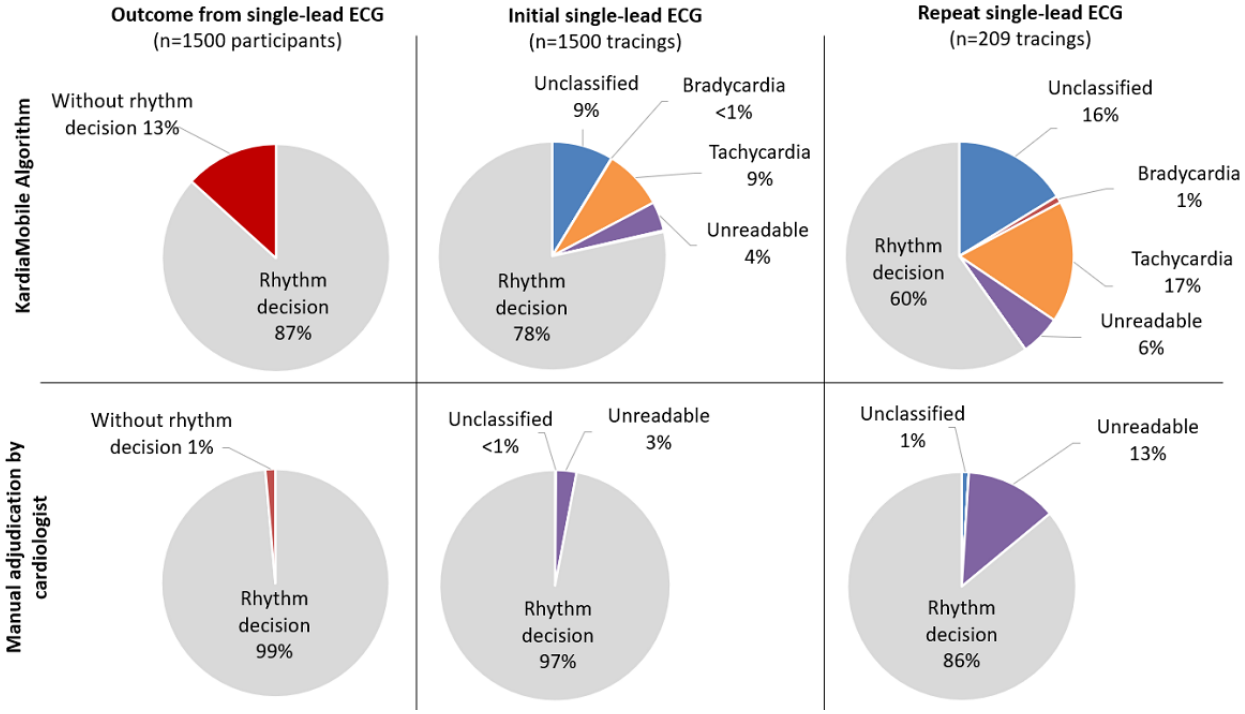


Table 2. KardiaMobile automated algorithm versus cardiologists' adjudication for single-lead electrocardiogram (ECG) for rhythm decision in n=1500 participants and atrial fibrillation detection in n=1709 ECG tracings.

KardiaMobile algorithm	Cardiologists' adjudication			
	Rhythm decision		Atrial fibrillation	
	Yes	No	Yes	No
Rhythm decision[a]				
Yes	1168	8	—[b]	—
No	287	37	—	—
Possible atrial fibrillation[c]				
Yes	—	—	9	61
No	—	—	3	1636

[a].

[b]Not applicable.

[c].

12-Lead ECG Analysis

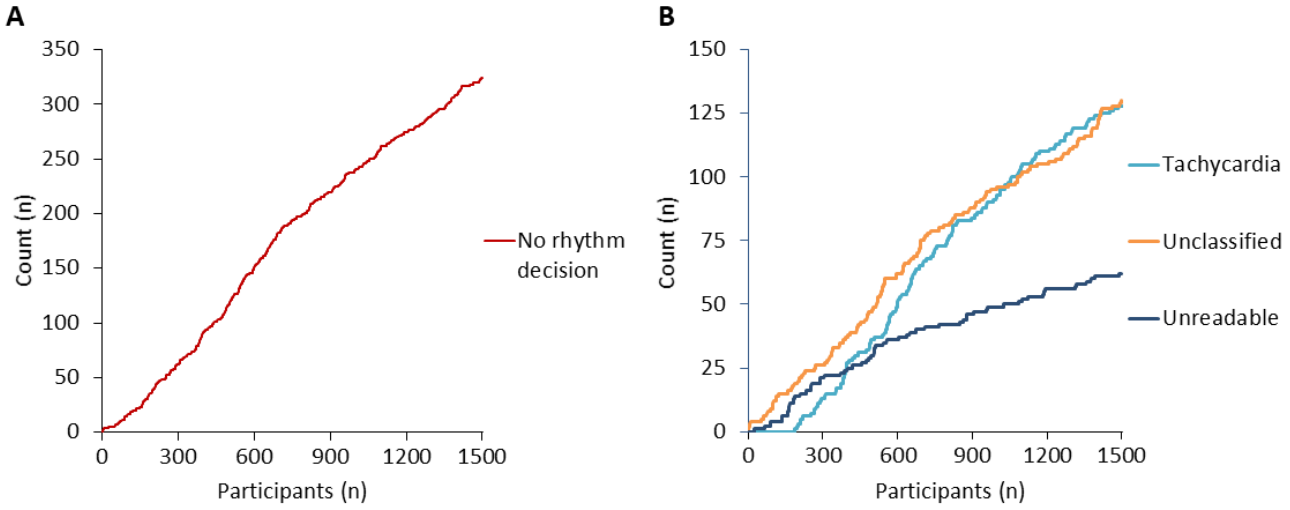
In total, 154 participants met criteria for a 12-lead ECG, but this was obtained in only 59.09% (91/154) due to participants not wanting to wait for the 12-lead ECG to be performed in the SCH emergency room. However, upon review of the single-lead ECGs meeting study criteria for a 12-lead ECG to be performed, the cardiologists adjudicated 89.61% (138/154) of these single lead ECGs to be of adequate quality for a rhythm decision. In total, diagnoses from the 12-lead ECGs were 89.01% (81/91) sinus rhythm, 1.10% (1/91) supraventricular tachycardia, and 9.89% (9/91) AF.

Utility of KardiaMobile Automated Algorithm for AF Screening

We analyzed the performance of the KardiaMobile automated algorithm for providing an initial rhythm decision. There was a linear relationship between ongoing participant recruitment and the occurrence of a no rhythm decision from the initial KardiaMobile tracing (Figure 5A). Linear regression analysis showed that there was a significant reduction in the cumulative incidence of no rhythm decision compared with successful rhythm decision with ongoing patient recruitment (β=–14.4, 95% CI –26.6 to –2.1; *P*=.02). As the KardiaMobile results of tachycardia, unclassified, and unreadable accounted for 98.77% (320/324) of occasions without a rhythm decision on the first KardiaMobile attempt, their contribution to no rhythm decision was further analyzed. With ongoing patient recruitment, the

occurrence of unreadable tracing was significantly reduced when compared with unclassified and tachycardia tracings (β=–38.0, 95% CI –63.3 to –12.6; *P*=.003, Figure 5B).

Figure 5. Cumulative occurrence and contributors to no rhythm decision from KardiaMobile's automated algorithm on initial electrocardiogram tracing: (A) cumulative occurrence of no rhythm decision from initial electrocardiogram tracing and (B) occurrence of unreadable tracing was significantly reduced when compared with unclassified and tachycardia tracings with increasing patient recruitment.



Discussion

Principal Findings

This study evaluated the utility of the KardiaMobile single-lead ECG device for AF screening in a semirural Ethiopian population of 1500 individuals from the TEFF-AF study. We found the KardiaMobile device performance to be suboptimal with successful automated rhythm decision following a single ECG trace of only 78%. This yield increased to 87% following a second KardiaMobile ECG tracing. As experience increased with ongoing patient recruitment, we encountered significant reduction in unreadable tracings. The ongoing occurrence of tachycardia and unclassified tracings contributed largely to the automated KardiaMobile algorithm's inability to achieve successful rhythm decision. In contrast, manual cardiologist assessment was able to obtain a rhythm decision in almost all cases (97%) with a single ECG. Taken together, our findings suggest that manual physician input remains necessary when the KardiaMobile device is used for AF screening.

The use of single-lead ECG devices is of increasing interest given the potential benefits of portability and scalability. Furthermore, automated rhythm analysis may allow for the use of such devices by individuals without formal medical training. However, there are limited data on the accuracy of these devices and their automated rhythm analysis algorithms in such settings despite the KardiaMobile device having been FDA-approved since 2012. In a small validation study, the KardiaMobile's automated AF detection algorithm was reported to yield high sensitivity of 98% and specificity of 97% with overall accuracy of 97% [13]. In a single-center, adjudicator-blinded case series of 52 consecutive patients with AF admitted for antiarrhythmic drug initiation who had serial 12-lead ECG and nearly simultaneously acquired KardiaMobile recordings, AF detection was reported at 96.6% sensitivity and 94.1% specificity [14]. However, 28% of the tracings obtained were deemed unclassified by the KardiaMobile automated algorithm and excluded from analysis. Similarly, others have reported the KardiaMobile automated algorithm correctly detected AF with 93% sensitivity and 84% specificity in 100 participants with a history of AF who presented for a scheduled elective electrical cardioversion after excluding a substantial 34% of recordings with unclassified tracings [11]. Our study found that the KardiaMobile automated algorithm failed to achieve rhythm decision in 22% of the tracings, comparable to previous studies. Consequently, this may limit the utility of the mobile single-lead ECG device for mass AF screening and opportunity to offer oral anticoagulation for stroke prevention in those with newly detected AF. It remains unclear if another mobile single-lead ECG device that was found to have higher sensitivity and similar specificity when compared with the KardiaMobile device will yield better AF screening performance [15].

Recently, several studies have reported on the use of other smart wearable devices using photoplethysmography-based technology for AF screening. The Apple Heart Study reported on the ability of a smartwatch photoplethysmography sensor and algorithm to screen individuals for an irregular pulse. Of 419,297 individuals, 2161 (0.52%) had a smartwatch-detected irregular pulse, with AF confirmed in 34% of those who returned an ECG patch. Of the individuals who had a smartwatch-detected

irregular pulse while simultaneously wearing an ECG patch, 84% (78/86) were in AF at the time [2]. The Huawei Heart Study similarly described the use of smartwatch or smartband photoplethysmography to screen 187,912 individuals. Of 227 with suspected AF who underwent complete history, examination, and ECG or 24-hour Holter monitoring, 87% were confirmed to have AF [3]. Although these data highlight the utility of automated algorithms to flag possible AF, both studies still incorporated physician review of confirmatory traces, and there remains a paucity of data comparing photoplethysmography-based and single-lead ECG technology.

Clinical Implications

Our study has important clinical implications for AF screening and highlights opportunities for future research. Prior research has shown that automated device algorithms can achieve accurate rhythm analysis under ideal conditions. However, our real-world experience in a resource-limited setting demonstrates that single-lead ECG tracing artefact and other limiting factors frequently prohibits algorithm interpretation. Despite limitations with tracing quality, manual cardiologist adjudication can still provide a rhythm diagnosis in the vast majority of cases. Thus, our findings suggest that physician input remains necessary for AF screening until further improvements in automated algorithms occur. In the meantime, repeat ECG tracings and increasing familiarity with using single-lead ECG devices are helpful to reduce unreadable tracings to improve diagnostic yield. Future studies should be undertaken to validate other mobile device technology and automated algorithms in real-world settings.

Limitations

Our screening protocol required a repeat tracing for occasions without a rhythm decision. However, this was not performed in a proportion of the participants, resulting in an incomplete data set. We acknowledge that the clinical value of AF screening in a young cohort with unknown risk factors for stroke is unclear. Nevertheless, given the knowns and unknowns of AF in sub-Saharan Africa and the higher prevalence of rheumatic heart disease, we did not restrict the AF screening to the typical target population of older individuals with higher stroke risk in developed countries [16]. As with all single time point AF screening, paroxysmal AF may be missed, leading to false negatives. Although our liberal inclusion criteria did achieve a diverse sample of the local community, we acknowledge that our data may not reflect the true prevalence of AF in this community due to the recruitment site being based at a local hospital.

Conclusion

The performance of the automated algorithm of the KardiaMobile single-lead ECG device was suboptimal when used for AF screening. However, the KardiaMobile device remains an excellent and affordable tool when used in low-resource settings with appropriate clinician input.

Acknowledgments

BP is supported by a Doctoral Scholarship from the Hospital Research Foundation. CW is supported by a Postdoctoral Fellowship from the National Heart Foundation of Australia. CW and DL are supported by Mid-Career Fellowships from the Hospital Research Foundation. PS is supported by a Practitioner Fellowship from the National Health and Medical Research Council of Australia and by the National Heart Foundation of Australia.

Authors' Contributions

All authors have full access to all the data and take full responsibility for the integrity of the data and the accuracy of data analysis. BP, SHC, AC, and DHL were responsible for study design and conception. BP, SHC, CXW, AJ, SI, GT, and AC were responsible for data acquisition and analysis. BP, CXW, AJ, SI, GT, PS, and DHL interpreted the data. BP, CXW, PS, and DHL drafted and revised the manuscript.

Conflicts of Interest

CXW reports that the University of Adelaide has received on his behalf lecture, travel, and/or research funding from Abbott Medical, Bayer, Boehringer Ingelheim, Medtronic, Novartis, Servier, and St Jude Medical. PS reports having served on the advisory board of Medtronic, Abbott Medical, Boston Scientific, Pacemate, and CathRx. PS reports that the University of Adelaide has received on his behalf lecture and/or consulting fees from Medtronic, Abbott Medical, and Boston Scientific. PS reports that the University of Adelaide has received on his behalf research funding from Medtronic, Abbott Medical, Boston Scientific, and MicroPort CRM. DHL reports that the University of Adelaide has received on his behalf lecture and/or consulting fees from Abbott Medical, Bayer, Boehringer Ingelheim, Biotronik, Medtronic, MicroPort CRM, and Pfizer.

References

1. Raja JM, Elsakr C, Roman S, Cave B, Pour-Ghaz I, Nanda A, et al. Apple Watch, wearables, and heart rhythm: where do we stand? Ann Transl Med 2019 Sep;7(17):417 [FREE Full text] [doi: 10.21037/atm.2019.06.79] [Medline: 31660316]
2. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]

3. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, MAFA II Investigators. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019 Nov 12;74(19):2365-2375. [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
4. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation 2014 Feb 25;129(8):837-847 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.005119] [Medline: 24345399]
5. Jonas DE, Kahwati LC, Yun JDY, Middleton JC, Coker-Schwimmer M, Asher GN. Screening for atrial fibrillation with electrocardiography: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2018 Aug 07;320(5):485-498. [doi: 10.1001/jama.2018.4190] [Medline: 30088015]
6. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
7. Soni A, Karna S, Fahey N, Sanghai S, Patel H, Raithatha S, et al. Age-and-sex stratified prevalence of atrial fibrillation in rural Western India: results of SMART-India, a population-based screening study. Int J Cardiol 2019 Apr 01;280:84-88 [FREE Full text] [doi: 10.1016/j.ijcard.2018.12.016] [Medline: 30551905]
8. Zaprutko T, Zaprutko J, Baszko A, Sawicka D, Szałek A, Dymecka M, et al. Feasibility of atrial fibrillation screening with mobile health technologies at pharmacies. J Cardiovasc Pharmacol Ther 2020 Mar;25(2):142-151. [doi: 10.1177/1074248419879089] [Medline: 31578088]
9. Rosenfeld LE, Amin AN, Hsu JC, Oxner A, Hills MT, Frankel DS. The Heart Rhythm Society/American College of Physicians atrial fibrillation screening and education initiative. Heart Rhythm 2019 Aug;16(8):e59-e65. [doi: 10.1016/j.hrthm.2019.04.007] [Medline: 30954599]
10. Lown M, Yue AM, Shah BN, Corbett SJ, Lewith G, Stuart B, et al. Screening for atrial fibrillation using economical and accurate technology (from the SAFETY Study). Am J Cardiol 2018 Oct 15;122(8):1339-1344. [doi: 10.1016/j.amjcard.2018.07.003] [Medline: 30131106]
11. William AD, Kanbour M, Callahan T, Bhargava M, Varma N, Rickard J, et al. Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: the iREAD study. Heart Rhythm 2018 Dec;15(10):1561-1565. [doi: 10.1016/j.hrthm.2018.06.037] [Medline: 30143448]
12. Wegner FK, Kochhäuser S, Ellermann C, Lange PS, Frommeyer G, Leitz P, et al. Prospective blinded Evaluation of the smartphone-based AliveCor Kardia ECG monitor for Atrial Fibrillation detection: the PEAK-AF study. Eur J Intern Med 2020 Mar;73:72-75. [doi: 10.1016/j.ejim.2019.11.018] [Medline: 31806411]
13. Lau JK, Lowres N, Neubeck L, Brieger DB, Sy RW, Galloway CD, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. Int J Cardiol 2013 Apr 30;165(1):193-194. [doi: 10.1016/j.ijcard.2013.01.220] [Medline: 23465249]
14. Bumgarner J, Lambert C, Cantillon D, Baranowski B, Wolski K, Hussein A, et al. Assessing the accuracy of an automated atrial fibrillation detection algorithm using novel smartwatch technology among patients presenting for elective cardioversion. J Am Coll Cardiol 2018 Mar;71(11):A411. [doi: 10.1016/s0735-1097(18)30952-5]
15. Desteghe L, Raymaekers Z, Lutin M, Vijgen J, Dilling-Boer D, Koopman P, et al. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. Europace 2017 Jan;19(1):29-39. [doi: 10.1093/europace/euw025] [Medline: 26893496]
16. Linz D, Sanders P, Pitman B, Dobrev D, Lau DH. Atrial fibrillation in sub-Saharan Africa: the knowns and unknowns. Int J Cardiol Heart Vasc 2019 Mar;22:212-213 [FREE Full text] [doi: 10.1016/j.ijcha.2019.02.008] [Medline: 30963098]

Abbreviations

AF: atrial fibrillation
ECG: electrocardiogram
FDA: US Food and Drug Administration
KM: KardiaMobile
SCH: Soddo Christian Hospital
TEFF-AF: The Heart of Ethiopia: Focus on Atrial Fibrillation study

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by B Stambler, G Giebel; comments to author 07.10.20; revised version received 21.11.20; accepted 12.12.20; published 19.05.21.

Please cite as:
Pitman BM, Chew SH, Wong CX, Jaghoori A, Iwai S, Thomas G, Chew A, Sanders P, Lau DH
Performance of a Mobile Single-Lead Electrocardiogram Technology for Atrial Fibrillation Screening in a Semirural African Population: Insights From "The Heart of Ethiopia: Focus on Atrial Fibrillation" (TEFF-AF) Study
JMIR Mhealth Uhealth 2021;9(5):e24470
URL: https://mhealth.jmir.org/2021/5/e24470
doi:10.2196/24470
PMID:34009129

Original Paper

The Use of SMS Text Messaging to Improve the Hospital-to-Community Transition in Patients With Acute Coronary Syndrome (Txt2Prevent): Results From a Pilot Randomized Controlled Trial

Emily S Ross[1], PhD; Brodie M Sakakibara[2,3], PhD; Martha H Mackay[4,5], RN, CCN(C), PhD; David G T Whitehurst[6,7], PhD; Joel Singer[5,8], PhD; Mustafa Toma[9], MD; Kitty K Corbett[6,10], PhD; Harriette G C Van Spall[11,12,13], MPH, MD; Kimberly Rutherford[14], MSc, MD; Bobby Gheorghiu[15], MHSc; Jillianne Code[16], PhD; Scott A Lear[1,6,17], PhD

[1]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada
[2]Centre for Chronic Disease Prevention and Management, University of British Columbia Okanagan, Kelowna, BC, Canada
[3]Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, BC, Canada
[4]School of Nursing, University of British Columbia, Vancouver, BC, Canada
[5]Centre for Health Evaluation and Outcome Sciences, Vancouver, BC, Canada
[6]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada
[7]Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada
[8]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[9]Division of Cardiology, Providence Health Care, Vancouver, BC, Canada
[10]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada
[11]Department of Medicine, McMaster University, Hamilton, ON, Canada
[12]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada
[13]Population Health Research Institute, Hamilton, ON, Canada
[14]Department of Family Practice, University of British Columbia, Vancouver, BC, Canada
[15]Canada Health Infoway, Toronto, ON, Canada
[16]Department of Curriculum and Pedagogy, University of British Columbia, Vancouver, BC, Canada
[17]Division of Cardiology, Providence Health Care, Healthy Heart Program St Paul's Hospital, Vancouver, BC, Canada

Corresponding Author:
Scott A Lear, PhD
Division of Cardiology
Providence Health Care
Healthy Heart Program St Paul's Hospital
B180-1081 Burrard Street
Vancouver, BC, V6Z 1Y6
Canada
Phone: 1 604 682 2344 ext 62778
Fax: 1 604 806 8590
Email: slear@providencehealth.bc.ca

Abstract

Background: Acute coronary syndrome (ACS) is a leading cause of hospital admission in North America. Many patients with ACS experience challenges after discharge that impact their clinical outcomes and psychosocial well-being. SMS text messaging has the potential to provide support to patients during this postdischarge period.

Objective: This study pilot tested a 60-day SMS text messaging intervention (Txt2Prevent) for patients with ACS. The primary objective was to compare self-management domains between usual care and usual care plus Txt2Prevent. The secondary objectives were to compare medication adherence, health-related quality of life, self-efficacy, and health care resource use between groups. The third objective was to assess the feasibility of the study protocol and the acceptability of the intervention.

Methods: This was a randomized controlled trial with blinding of outcome assessors. We recruited 76 patients with ACS from St. Paul's Hospital in Vancouver, Canada, and randomized them to 1 of 2 groups within 7 days of discharge. The Txt2Prevent program included automated 1-way SMS text messages about follow-up care, self-management, and healthy living. Data were collected during the index admission and at 60 days after randomization. The primary outcome was measured with the Health Education Impact Questionnaire (heiQ). Other outcomes included the EQ-5D-5L, EQ-5D-5L Visual Analog Scale, a modified Sullivan Cardiac Self-Efficacy Scale, and Morisky Medication Adherence Scale scores, and self-reported health care resource use. Analyses of covariance were used to test the effect of group assignment on follow-up scores (controlling for baseline) and were considered exploratory in nature. Feasibility was assessed with descriptive characteristics of the study protocol. Acceptability was assessed with 2 survey questions and semistructured interviews.

Results: There were no statistically significant differences between the groups for the heiQ domains (adjusted mean difference [Txt2Prevent minus usual care] for each domain—Health-directed activity: –0.13, 95% CI –0.39 to 0.13, *P*=.31; Positive and active engagement in life: 0.03, 95% CI –0.19 to 0.25, *P*=.76; Emotional distress: 0.04, 95% CI –0.22 to 0.29, *P*=.77; Self-monitoring and insight: –0.14, 95% CI –0.33 to 0.05, *P*=.15; Constructive attitudes and approaches: –0.10, 95% CI –0.36 to 0.17, *P*=.47; Skill technique and acquisition: 0.05, 95% CI –0.18 to 0.27, *P*=.69; Social integration and support: –0.12, 95% CI –0.34 to 0.10, *P*=.27; and Health services navigation: –0.05, 95% CI –0.29 to 0.19, *P*=.69). For the secondary outcomes, there were no statistically significant differences in adjusted analyses except in 1 self-efficacy domain (Total plus), where the Txt2Prevent group had lower scores (mean difference –0.36, 95% CI –0.66 to –0.50, *P*=.03). The study protocol was feasible, but recruitment took longer than expected. Over 90% (29/31 [94%]) of participants reported they were satisfied with the program.

Conclusions: The Txt2Prevent study was feasible to implement; however, although exploratory, there were no differences between the 2 groups in adjusted analyses except for 1 self-efficacy domain. As the intervention appeared acceptable, there is potential in using SMS text messages in this context. The design of the intervention may need to be reconsidered to have more impact on outcome measures.

Trial Registration: ClinicalTrials.gov NCT02336919; https://clinicaltrials.gov/ct2/show/NCT02336919

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.6968

(JMIR Mhealth Uhealth 2021;9(5):e24530) doi:10.2196/24530

KEYWORDS

SMS text messaging; mHealth; acute coronary syndrome; cardiovascular disease

Introduction

Acute coronary syndrome (ACS), which includes unstable angina and myocardial infarction, is a leading cause of hospitalization in North America [1,2]. Once discharged, 20%-34% of patients are readmitted within 30 days [3,4]. While reducing readmissions is a complex issue, patients with ACS may experience several challenges after discharge that negatively impact their clinical outcomes and psychosocial well-being. One-third of patients with ACS do not adhere to the behavioral advice regarding diet, physical activity, and smoking cessation [5]. In Canada, 44% of patients with myocardial infarction do not have early physician follow-up within 7 days, particularly those who live in rural areas and lower-income neighborhoods [6]. Physician follow-up may be important to reduce readmissions and improve medication adherence [7-10]. Additionally, around a quarter of cardiac medication prescriptions are not filled within the first week of discharge [11]. Patients also report that once they return home they can feel overwhelmed or uncertain [12], be fearful of another cardiac event [13], and experience depression [12,14]. Their time in the hospital can be busy and overwhelming, and as such, some patients may have difficulty remembering everything they are told or may not know what to ask [12,15]. Furthermore, patients' length of stay in the hospital has markedly declined over the past several decades [16,17], which means there can be less time to deliver patient education. Therefore, providing continuing support after hospital discharge may affect several key factors of post-ACS management, including lifestyle changes, medication adherence, and psychosocial well-being.

Home-based programs, often nurse led, can improve quality of life and reduce readmissions [18,19], but these face-to-face interventions can be a challenge for strained health care systems. The widespread use of information and communication technology, such as mobile phones, may be an easier and more convenient way to reach patients. SMS text messages are an attractive technology, as over 90% of adults aged 65 years or older own a cell phone [20], and 80% of cell phone owners currently text [21]. SMS text messages also have the benefits of being able to store messages that can be reaccessed, have a wide geographic reach, are convenient due to the asynchronous nature of communication, and have low delivery costs. Previous SMS text messaging studies in patients with or at risk for cardiovascular disease (CVD) have reported improvements in self-management behaviors (eg, medication adherence [22] and increases in leisure physical activity and walking [23]) and cardiac risk factors (eg, lowering low-density lipoprotein cholesterol and systolic blood pressure [24,25]). These studies show the promise of using SMS text messaging to aid in the care of patients with CVD. However, they do not specifically target the multiple self-management behaviors required in the immediate period after discharge using only SMS text messages. We report on a pilot study of a one-way SMS text messaging intervention (Txt2Prevent) aimed at supporting patients with ACS after hospital discharge in an assessor-blinded randomized

controlled trial. The study's primary objective was to assess the effect of the Txt2Prevent intervention on self-management domains compared with usual care. The other objectives were to compare quality of life, self-efficacy, medication adherence, and health care resource use between the 2 groups as well as to assess the feasibility of the study protocol and acceptability of the Txt2Prevent intervention.

Methods

Study Design

The Txt2Prevent study is a mixed method, assessor-blinded, randomized controlled trial with a parallel group design. The study protocol and intervention development have been previously reported [26]. This study is reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) eHealth checklist [27] and is registered at ClinicalTrials.gov (NCT02336919).

Participants

Patients with a diagnosis of ACS, as identified by clinical staff, were recruited from St. Paul's Hospital, a tertiary care hospital, in Vancouver, Canada between June 2015 and October 2016. Patients were eligible to participate if they had ACS (unstable angina or any type of myocardial infarction) as their primary admitting diagnosis, had daily access to a phone with SMS text messaging capabilities, were able to provide informed consent, and were able to read and understand English. Exclusion criteria comprised having coronary artery bypass graft surgery as a treatment for the ACS admission, having a prescheduled surgery within the study period, an expectation that the individual would not survive the duration of the study due to non-CVD reasons, being discharged to a long-term care center, or living outside the province of British Columbia. As this was a pilot study, the sample size was based on convenience. We aimed to recruit 76 participants as we previously estimated this was feasible over 6 months of recruitment. All participants provided written informed consent. Ethics and institutional approvals were obtained from Providence Health Care Research Ethics Board and Simon Fraser University's Office of Research Ethics.

Baseline questionnaires, which included demographic information as well as measures of self-management, health-related quality of life, and cardiac self-efficacy, were administered in-person in the hospital when possible or within 7 days after discharge. Clinical information was gathered from the participant's medical record.

Randomization

After participants completed the baseline questionnaires and were discharged from the hospital, they were randomly assigned to either the intervention (Txt2Prevent plus usual care) or usual care. A statistician not associated with the study generated a random allocation schedule, which randomized participants in a 1:1 ratio using variable block sizes, stratified by sex. A research assistant not involved in recruitment or outcome assessment accessed a secure randomization database to obtain allocations for each patient and informed participants of their group assignment.

Intervention and Usual Care

Participants in the Txt2Prevent group received 48 unique, automated, one-way messages over 60 days following randomization in addition to usual care. An additional 4 messages relating to study administration (eg, indicating the end of the study and requesting participants to inform us if they were readmitted; see SMS text messages sent on days 7, 26, 45, and 60 in Multimedia Appendix 1) were also sent during the study period. The SMS text messages were delivered at a time of day specified by the participant, began after the participant was randomized, and were sent daily for the first 36 days and then every other day until day 60. We had received feedback from patients in the design stage that it would be helpful to have more support in the beginning, but that daily messages may not be required for the whole period. We primarily covered time-sensitive information about recovery in the first 36 days. Once we covered the primary recovery topics suggested by our clinical advisory committee, we then switched to focus more on healthy living SMS text messages for the remainder of the study period. We chose 60 days for the program length because the initial period after discharge is the highest risk for readmission [4,28]. In addition, many patients will have seen their cardiologist or started a cardiac rehabilitation program by then [29] as well as may be starting to adjust to their new normal or have returned to work [30,31]. SMS text messages covered a range of topics, from time-sensitive information regarding their recovery (eg, timely follow-up with their health care professional) to general healthy living advice (eg, SMS text messages regarding physical activity, diet, and psychosocial health), and were delivered in a prespecified order (Table 1 and Multimedia Appendix 1). Participants received different SMS text messages on 2 occasions based on their smoking status; no other aspects were personalized. The usual care group did not receive any SMS text messages or have any contact from research staff during the study period.

Table 1. Examples of the SMS text messages in the intervention group (Txt2Prevent).

Topic	Example SMS text message
Appointment reminders	T2P: See a heart specialist (a cardiologist or internist) within 6 weeks of discharge. If this isn't set up, call their office, or your family doctor. (Day 15)
Smoking cessation	T2P: Not smoking is one of the most important things you can do for your health. For quitting resources, check out: http://bit.ly/quitnowbc (Day 8)
Recovery guidelines	T2P: Resuming sex: A general guide is that if you can go up a flight of stairs without symptoms, it is probably safe to restart sexual activities. (Day 14)
Psychosocial	T2P: It is common to feel sad or depressed after a heart attack or being in the hospital. If you feel this way for 2+ weeks, contact your doctor. (Day 16)
Physical activity	T2P: Have you done something physically active today? If you have questions, call the Physical Activity Line at 1-877-725-1149 or talk to your doctor. (Day 21)
Medication	T2P: Bring a list of your medications to your appointment when you see your doctor. You can get copies from your pharmacist. (Day 9)

Outcome Measures and Data Collection

The primary outcome was follow-up scores (controlled for baseline scores) in self-management domains as measured by the Health Education Impact Questionnaire (heiQ; version 3) [32]. The heiQ comprises 40 questions that cover 8 domains in total. All domains were measured and reported on separately: Health-directed behavior (4 questions), Positive and active engagement in life (5 questions), Emotional distress (6 questions), Self-monitoring and insight (6 questions), Constructive attitudes and approaches (5 questions), Skill and technique acquisition (4 questions), Social integration and support (5 questions), and Health service navigation (5 questions). As per the questionnaire's scoring instructions, each domain score was calculated by averaging Likert scale responses (scaled from 1 to 4). Higher values are desirable, except for the Emotional distress domain. The heiQ was developed using item response theory and structural equation modeling, and the subscales have "acceptable" to "high" internal consistency (Cronbach α ranging from .70 to .86, depending on the domain) [32].

The secondary outcomes were health-related quality of life, cardiac self-efficacy, medication adherence, and health care resource use. Health-related quality of life was measured with the EQ-5D-5L [33], using health state valuations derived from a representative sample of the Canadian adult general population [34]. Self-reported health status was also captured using the EQ-5D-5L Visual Analog Scale (EQ VAS), a 0-100 VAS with anchors defined as "the best health you can imagine" (100) and "the worst health you can imagine" (0) [35]. The EQ-5D has been validated and used around the world. We used the 5-level scale as it has higher discriminatory power, interobserver reliability, and test–retest reliability than the 3-level EQ-5D [33,36]. The 3-level scale was validated in an ACS population, and had high usability, reasonable criterion validity when compared with other quality of life scales, and good test–retest reliability [37]. Cardiac self-efficacy was measured with a modified Sullivan Cardiac Self-Efficacy scale (CSE) [38], such that scores were calculated for the 2 domains (Control symptoms and Maintain function) as well as for the total by averaging Likert scale responses (0-4). For the modifications, we combined the first 4 questions regarding symptom control into 2 questions, as well as added 3 questions about diet and emotional well-being. We calculated the total for the original questions (Total) as well as a total including the additional questions (Total plus). The original Sullivan CSE scale has high internal consistency (Cronbach α of .90 and .87 for the 2 scales, respectively), and good convergent and discriminant validity when compared with other distress and disability scales [38]. For our modifications, Cronbach α were .71, .76, .79, and .82 for "Control symptoms," "Maintain function," "Total," and "Total plus," respectively. Medication adherence was measured with the Morisky Medication Adherence Scale (MMAS-8) [39-41]. As per questionnaire documentation, we calculated an adherence score on a scale from 1 to 8 and categorized participants as having low (<6), medium (6 to <8), or high adherence (8). This 8-item medication adherence scale has good internal consistency (Cronbach α=.83) and reliability when assessed in a hypertensive population [39]. It has good sensitivity (93%) and moderate specificity (53%) [39]. We also assessed how many participants at follow-up reported taking the recommended medications for post-ACS treatment [42]: acetylsalicylic acid, ticagrelor/clopidogrel, a statin, a beta blocker, and an angiotensin-converting-enzyme inhibitor/angiotensin receptor blocker for those with reduced vascular function. Health care resource use over the 60-day follow-up period (eg, visits to health care practitioners, visits to hospitals, cardiac rehabilitation program participation) was self-reported through a questionnaire developed by the research team. Any self-reported hospitalizations were verified with hospital records. Two blinded assessors (MM and MT) categorized hospital readmissions as cardiac or noncardiac.

Study feasibility was assessed through descriptive statistics on recruitment rates, follow-up rates, questionnaire completion rates, method of questionnaire completion (eg, postal mail, phone), percentage of participants randomized within 7 days, and percentage of participants who completed follow-up within 6 weeks after finishing the 60-day study period. In addition, study staff kept a log of barriers encountered. Acceptability was measured via two 5-level Likert scale survey questions that asked how satisfied participants were with the program (strongly disagree to strongly agree) and whether they thought the program helped them manage their condition. Acceptability was also assessed via 2 questions in semistructured phone

interviews—specifically whether they would recommend the program to other patients with ACS and whether they read the SMS text messages. Participants with a range of demographic characteristics who were randomized to the Txt2Prevent group were invited to participate in the semistructured interviews after the 60-day study period. Detailed findings from the interviews, which covered participants' experiences with the program and gathered feedback on program attributes, will be presented in a separate paper.

Follow-up questionnaires were administered 60 days after randomization primarily via postal mail, except for the health care resource use questions, which were completed over the phone for most participants due to its complex branching. All surveys were administered at baseline and at follow-up, except for the medication adherence scale and health care resource use questionnaire, which were only administered at follow-up.

Statistics

As this was a pilot study, we undertook similar analyses to what is anticipated for a full trial and considered them exploratory in nature. Descriptive statistics are presented as the mean and SD, or count data with percentages. Analyses were conducted following the intent-to-treat principle. Only complete cases were analyzed. For continuous data from the heiQ, EQ-5D-5L, EQ VAS, and CSE questionnaires, analysis of covariance (ANCOVA) was used to test the effect of group assignment on the follow-up scores when controlling for baseline scores. We then reran the ANCOVA adjusting for age and sex as prespecified covariates as well as previous CVD status for the heiQ, and previous CVD status and marital status for the cardiac self-efficacy, due to their prognostic value [43]. We tested for the following assumptions for ANCOVA tests: independence of covariate and treatment effect, homogeneity of regression slopes, linearity, normality of residuals, homoscedasticity, and homogeneity of variances. The CSE scales, the EQ-5D-5L, and the EQ VAS were negatively skewed, which was primarily driven by outliers. We conducted the analyses with and without the outliers and present results of both analyses in situations where the outlier impacted the conclusion. We also calculated Cohen *d* effect sizes (mean difference score from the Txt2Prevent group minus mean difference score for the usual care group divided by pretest pooled SD [44]) to provide more context for the continuous data results. For our questionnaires, a negative Cohen *d* effect size implies the Txt2Prevent group had worse outcomes than the usual care group.

For count data from the health care resource use questionnaire (eg, number of readmissions), we used negative binomial regression analyses as our data had some overdispersion. For these analyses, we adjusted for age and sex, as prespecified. For binary response data from the health care resource use questionnaire (eg, cardiologist visit within 60 days—yes/no), we used a robust Poisson regression to determine relative risk as our outcomes occurred frequently [45]. For the robust Poisson regression, we adjusted for age and sex as prespecified, as well as geographic region [6] and income [6]. Chi-square tests were used for categorical variables unless there were low expected counts, in which case the Fisher exact test was used. Analyses were performed using SPSS (version 25; IMB Corp). Statistical significance was set at $P<.05$.

Results

Participant Demographics

Four hundred patients were assessed for eligibility from June 2015 to October 2016. After excluding those who did not meet inclusion criteria and those who declined to participate, 76 participants were randomized (Figure 1).

The mean age of participants was 60 years (SD 9.3) and 73% (55/75) were male (Tables 2 and 3). Nine participants did not complete the study (2 withdrew, 3 failed to complete any of the follow-up questionnaires, and 4 partially completed the follow-up questionnaires; Figure 1). There were no statistically significant differences in baseline characteristics between those who did not complete the follow-up assessments compared with those who did, including in age (t_{73}=–0.16 [2 tailed], P=.87), sex (Fisher exact test, P=.18), and whether or not they completed education past high-school (Fisher exact test, P=.09; full analyses not shown). Data collection ended in December 2016.

Figure 1. CONSORT flow diagram.

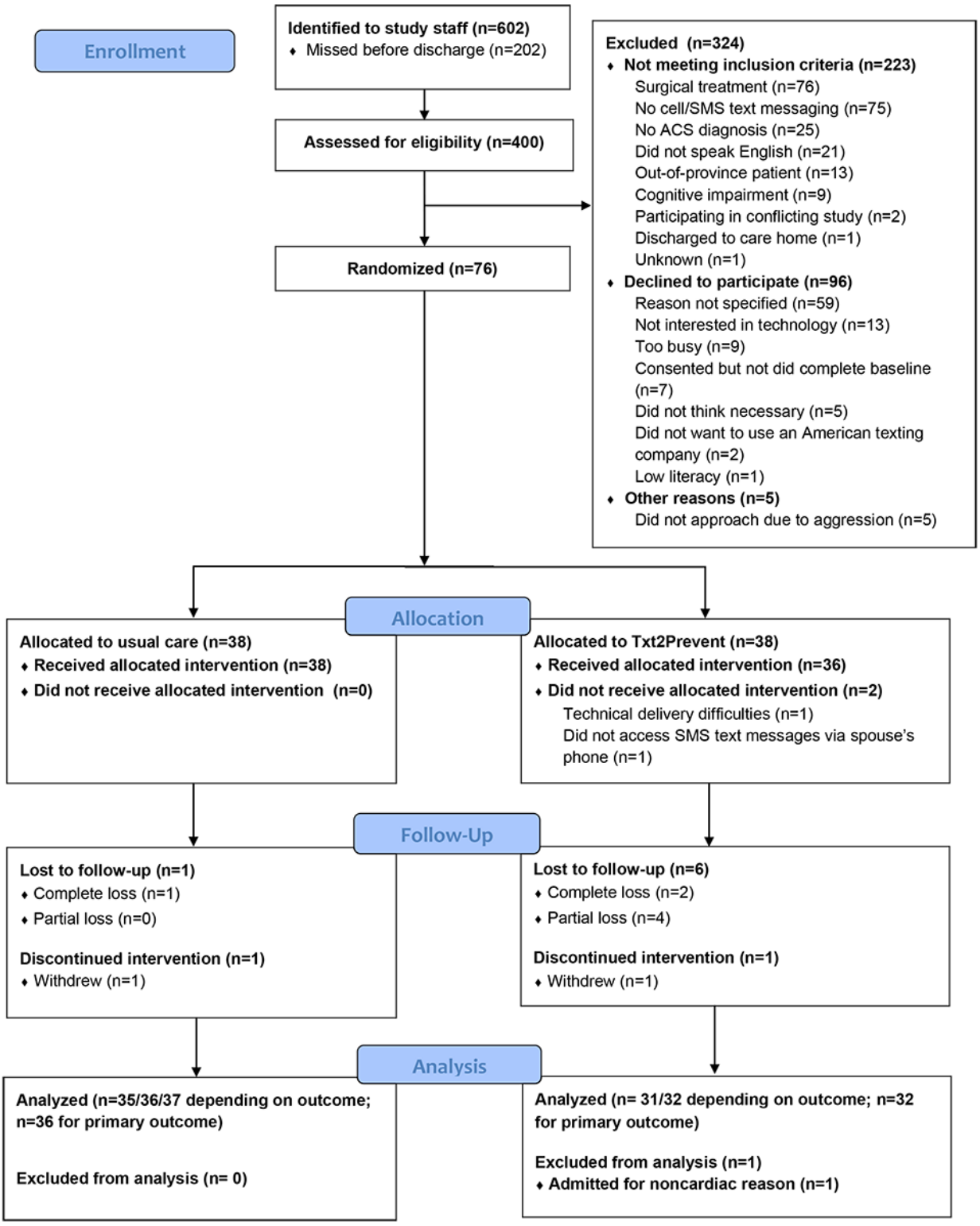


Table 2. Baseline demographics, by group.

Variable	Group[a]	
	Txt2Prevent	Usual care
Socioeconomic status		
Age, mean (SD)	59.5 (9.1)	61.1 (9.6)
Male, n (%)	27 (73)	28 (74)
Married (yes/no), n (%)	27 (73)	31 (82)
Geographic region, n (%)		
Census metropolitan area (100,000+ urban core) [46]	13 (35)	21 (55)
Census agglomeration (10,000-99,999 urban core) [46]	18 (49)	8 (21)
Rural [46]	6 (16)	9 (24)
Greater than high-school education (yes/no), n (%)	23 (62)	25 (66)
Employed full-time (yes/no), n (%)	20 (54)	17 (45)
Household income, n/N (%)[b]		
Less than Can $29,999[a]	6/33 (18)	7/36 (19)
Can $30,000 to Can $69,999	9/33 (27)	7/36 (19)
Can $70,000 to Can $99,999	6/33 (18)	6/36 (17)
Can $100,000 or higher	12/33 (36)	16/36 (44)
Technology use, n/N (%)[b]		
At least daily cell use	26/36 (72)	34/38 (89)
Very or completely confident using a cell phone	18/33 (55)	26/37 (70)
Own a smartphone	34/37 (92)	33/37 (89)
Comorbidities and medical history, n (%)		
Hypertension	25 (68)	19 (50)
Dyslipidemia	18 (49)	12 (32)
Diabetes (type 1 or type 2)	14 (38)	7 (18)
Previous any type of cardiovascular disease	16 (43)	16 (42)
Treatment in hospital, mean (SD)		
Days in hospital	5.1 (3.0)	5.2 (4.2)
Primary reason for admission, n (%)		
Non-ST-segment elevation acute coronary syndrome	10 (27)	18 (47)
ST-segment elevation myocardial infarction	22 (59)	17 (45)
Other	5 (14)	3 (8)
Revascularization, n (%)	29 (78)	33/37 (89)
Current/quit within 6-month smoker, n (%)	8 (22)	8/37 (22)
Medication at discharge, n/N (%)		
Acetylsalicylic acid	33/36 (92)	36/37 (97)
Ticagrelor or clopidogrel	31/36 (86)	36/37 (97)
Statin	33/36 (92)	33/37 (89)
Beta blocker	29/36 (81)	33/37 (89)
Angiotensin-converting-enzyme inhibitor or angiotensin II receptor blocker	29/36 (81)	32/37 (86)

[a]n=37 and =38 for the Txt2Prevent and Usual care groups unless stated otherwise.

[b]Can $1 = US $0.81.

Table 3. Baseline questionnaire scores, by group.

Baseline variables, mean (SD)	Group	
	Txt2Prevent (n=37)	Usual care (n=38)
Health Education Impact Questionnaire (heiQ)		
Health-directed activity	2.93 (0.80)	2.93 (0.78)
Positive and active engagement in life	3.13 (0.53)	3.27 (0.48)
Emotional distress	2.25 (0.68)	2.02 (0.60)
Self-monitoring and insight	3.19 (0.89)	3.08 (0.59)
Constructive attitudes and approaches	3.17 (0.51)	3.35 (0.46)
Skill technique and acquisition	3.02 (0.34)	3.09 (0.53)
Social integration and support	3.11 (0.49)	3.30 (0.41)
Health service navigation	3.08 (0.47)	3.27 (0.47)
EQ-5D-5L	0.833 (0.119)	0.849 (0.109)
EQ-5D-5L Visual Analog Scale (EQ VAS)	67.00 (19.00)	68.00 (17.00)
Cardiac Self-Efficacy (CSE)		
Symptoms	3.14 (0.63)	3.25 (0.52)
Function	2.85 (0.91)	2.91 (0.83)
Total	3.02 (0.61)	3.10 (0.57)
Total plus	2.95 (0.58)	3.02 (0.54)

Primary Outcome

There were no statistically significant differences between groups for the heiQ scores in either the unadjusted (Multimedia Appendix 2) or adjusted model in any of the 8 domains (adjusted mean difference [Txt2Prevent minus usual care] for each domain: Health-directed activity: –0.13, 95% CI –0.39 to 0.13, *P*=.31; Positive and active engagement in life: 0.03, 95% CI –0.19 to 0.25, *P*=.76; Emotional distress: 0.04, 95% CI –0.22 to 0.29, *P*=.77; Self-monitoring and insight: –0.14, 95% CI –0.33 to 0.05, *P*=.15; Constructive attitudes and approaches: –0.10, 95% CI –0.36 to 0.17, *P*=.47; Skill technique and acquisition: 0.05, 95% CI –0.18 to 0.27, *P*=.69; Social integration and support: –0.12, 95% CI –0.34 to 0.10, *P*=.27; Health services navigation: –0.05, 95% CI –0.29 to 0.19, *P*=.69) (Table 4). Cohen *d* effect sizes were all below 0.20, indicating negligible effects, except for the Self-monitoring and insight domain, where the Txt2Prevent group had worse outcomes than the usual care group, estimated at a small negative effect (*d*=–0.48; Table 4).

Table 4. Adjusted 60-day Health Education Impact Questionnaire (heiQ) scores, by group.[a]

Outcome	Group		Adjusted mean difference (95% CI)	*P* value	Effect size (Cohen *d*)[b]
	Txt2Prevent (n=32), adjusted mean (95% CI)	Usual care (n=36), adjusted mean (95% CI)			
Health Education Impact Questionnaire (heiQ)					
Health-directed activity[c]	3.02 (2.82 to 3.21)	3.15 (2.96 to 3.35)	–0.13 (–0.39 to 0.13)	.31	–0.15
Positive and active engagement in life	3.10 (2.93 to 3.26)	3.06 (2.91 to 3.22)	0.03 (–0.19 to 0.25)	.76	0.10
Emotional distress	2.37 (2.18 to 2.56)	2.33 (2.15 to 2.51)	0.04 (–0.22 to 0.29)	.77	–0.05
Self-monitoring and insight	3.08 (2.94 to 3.23)	3.22 (3.09 to 3.36)	–0.14 (–0.33 to 0.05)	.15	–0.48
Constructive attitudes and approaches	3.09 (2.89 to 3.29)	3.18 (2.99 to 3.38)	–0.10 (–0.36 to 0.17)	.47	–0.06
Skill technique and acquisition	2.91 (2.73 to 3.08)	2.86 (2.70 to 3.03)	0.05 (–0.18 to 0.27)	.69	0.14
Social integration and support	3.04 (2.87 to 3.21)	3.17 (3.01 to 3.33)	–0.12 (–0.34 to 0.10)	.27	–0.04
Health services navigation	3.15 (2.97 to 3.33)	3.19 (3.02 to 3.37)	–0.05 (–0.29 to 0.19)	.69	0.15

[a]The adjusted heiQ model includes baseline scores, age, sex, and previous cardiovascular disease status (yes/no).

[b]Effect size is Cohen *d* (mean difference score from the Txt2Prevent group minus mean difference score for the usual care group divided by pretest pooled SD [44]). For our questionnaires, a negative number implies the Txt2Prevent group had worse outcomes than the usual care group.

[c]In the usual care group, 35 participants were analyzed for the Health-directed activity domain.

Secondary Outcomes

There were no statistically significant differences in EQ-5D-5L health state values (*P*=.51) or EQ VAS scores (*P*=.71; Table 5). For cardiac self-efficacy, in the adjusted models, there were no statistically significant differences between the 2 groups, except for the "Total plus" domain, where the Txt2Prevent group had worse outcomes (Control symptoms: *P*=.10; Maintain function: *P*=.05; Total: *P*=.05; Total plus: *P*=.03; Table 5). The statistically significant finding on the "Total plus" scale was due to an influential outlier that impacted the normality assumptions of ANCOVA. When the influential outlier was excluded in the adjusted analysis, the *P* value for the Total plus scale was no longer significant (*P*=.05). Depending on the self-efficacy domain, there were small or medium negative effects for the Txt2Prevent group (ie, this group had worse outcomes) based on the Cohen *d* values for the self-efficacy scores (Table 5).

Table 5. Adjusted 60-day EQ-5D-5L Visual Analog Scale (EQ VAS), EQ-5D-5L, and Cardiac Self-Efficacy results, by group.[a]

Outcome	Group		Adjusted mean difference (95% CI)	Adjusted *P* value	Effect size (Cohen *d*)[b]
	Txt2Prevent (n=32), adjusted mean (95% CI)	Usual care (n=36), adjusted mean (95% CI)			
EQ-5D-5L Visual Analog Scale (EQ VAS)	70.94 (65.91 to 75.98)	69.68 (64.84 to 74.52)	–1.27 (–5.41 to 7.94)	.71	0.10
EQ-5D-5L[c]	0.82 (0.78 to 0.86)	0.84 (0.80 to 0.88)	–0.018 (–0.07 to 0.04)	.51	–0.13
Cardiac Self-Efficacy Scale (CSE)					
Control symptoms	2.49 (2.24 to 2.75)	2.76 (2.49 to 3.02)	–0.27 (–0.58 to 0.05)	.10	–0.43
Control symptoms (2 outliers removed)[c]	2.57 (2.36 to 2.78)	2.80 (2.57 to 3.02)	–0.23 (–0.49 to 0.04)	.09	–0.37
Maintain function	2.14 (1.84 to 2.45)	2.52 (2.20 to 2.84)	–0.38 (–0.76 to 0.004)	.05	–0.46
Maintain function (1 outlier removed)[c]	2.23 (1.95 to 2.50)	2.50 (2.22 to 2.78)	–0.27 (–0.61 to 0.07)	.11	–0.35
Total	2.35 (2.09 to 2.60)	2.66 (2.39 to 2.93)	–0.31 (–0.63 to 0.003)	.05	–0.55
Total (1 outlier removed)[c]	2.42 (2.20 to 2.64)	2.64 (2.41 to 2.86)	–0.22 (–0.49 to 0.05)	.11	–0.40
Total plus	2.28 (2.03 to 2.53)	2.64 (2.38 to 2.90)	–0.36 (–0.66 to –0.05)	.03	–0.65
Total plus (1 outlier removed)[c]	2.35 (2.14 to 2.57)	2.62 (2.39 to 2.84)	–0.26 (–0.53 to 0.003)	.05	–0.51

[a]The adjusted EQ-5D-5L and EQ VAS models include baseline scores, age, and sex. The adjusted CSE model includes baseline scores, age, sex, marital status, and previous cardiovascular disease status (yes/no).

[b]Effect size is Cohen *d* (mean difference score from the Txt2Prevent group minus mean difference score for the usual care group divided by pretest pooled SD [44]). For our questionnaires, a negative number implies the Txt2Prevent group had worse outcomes than the usual care group.

[c]In the Txt2Prevent group, 31 participants were analyzed for the EQ-5D-5L questionnaire. Thirty-two participants were analyzed for the remaining outcomes (excluding those with outliers removed). For the 2 outliers in the Control symptoms domain, 1 was from the Txt2Prevent group and 1 from the usual care group. The 1 outlier for the Maintain function, Total, and Total plus was from the Txt2Prevent group.

There were no statistically significant differences in the mean medication adherence scores between the 2 groups (*P*=.27; Multimedia Appendix 2). When categorized into low, medium, and high adherence, 34% (11/32) of those in the Txt2Prevent group and 42% (15/36) of those the usual care group were classified as high adherers (χ^2_2=2.10, *P*=.35). There were no statistically significant differences between the 2 groups for the categories of cardiac medications they were prescribed (Fisher exact test; *P* values ranged from .24 for statins to >.99 for beta blockers; Multimedia Appendix 3).

There were no differences between the groups in either the percentage of participants who visited the hospital or the mean number of visits to the hospital for all-cause or cardiac visits (Table 6). There were no differences in whether participants had visited a family physician, joined a cardiac rehabilitation program, or visited a cardiologist over the study period in adjusted analyses (Multimedia Appendix 4).

Table 6. Type of and mean hospital visits within 60 days, by group.

Outcome	Group[a]		*P* value	Group[b]		*P* value
	Txt2Prevent (n=32), adjusted mean visits (95% CI)	Usual care (n=37), adjusted mean visits (95% CI)		Txt2Prevent (n=32), n (%)	Usual care (n=37), n (%)	
Cardiac emergency department	0.00 (–)	0.08 (0.02-0.38)	N/A	0 (0)	3 (8)	.24
All-cause emergency department	0.20 (0.08-0.48)	0.33 (0.16-0.66)	.36	3 (9)	9 (24)	.10
Cardiac hospitalization	0.13 (0.04-0.37)	0.12 (0.04-0.34)	.92	3 (9)	3 (8)	1.00
All-cause hospitalization	0.16 (0.06-0.42)	0.21 (0.09-0.48)	.70	4 (13)	6 (16)	.74

[a]Mean visits were analyzed with a negative binomial regression adjusted for age and sex.

[b]Number of participants—n (%)—admitted for all-cause emergency department visits was analyzed with chi-square (*df*=1) while the remaining visit types were analyzed with a Fisher exact test due to low expected cell counts.

Assessment of Feasibility and Acceptability

Recruitment of the target sample took 17 months (Table 7), which was longer than the anticipated 6 months. As much as 55.8% (223/400) of patients we approached were ineligible (Figure 1), of which 34.1% (76/223) were scheduled for surgery and 33.6% (75/223) did not own a cell phone. Of those eligible, 55.8% (96/172) declined to participate (Figure 1). Our randomization system worked well as 97% (74/76) of participants were randomized within our target of 7 days of discharge and 66% (50/76) were randomized within 2 days (Table 7). Those that were not randomized within the target time frame were due to delayed completion of baseline questionnaires. Almost 91% of participants (69/76) completed follow-up for the primary outcome. We obtained all follow-up questionnaires for 88% (67/76) of participants and at least one follow-up questionnaire for 93% (71/76) of participants; however, one participant was excluded from the analysis because they were admitted for a noncardiac reason. While we developed electronic versions to provide an alternative option, the majority of participants were willing to complete the questionnaires in their default format (60/69 [87%] completed packaged questionnaires by postal mail and 65/69 [94%] completed the health care resource use by telephone; Table 7).

We had a technical problem with our delivery system 11 months into recruitment where 80 SMS text messages were not delivered for 10 days for 28% (10/36) participants. It is suspected an operating system update caused the error as a server reboot fixed the error. Once we fixed the error, all affected participants resumed the SMS text messages where they left off. In order to keep blinding and consistency in the timing of the outcome measurements, follow-up assessments were still scheduled for 60-days after randomization. Of the 10 affected participants, 2 did not complete follow-up, 5 completed the primary outcome assessment between 60 and 70 days after randomization, and 3 completed follow-up after 70 days. After this technical problem, we implemented more regular system checks by the staff involved in randomization to ensure all SMS text messages were being delivered.

Regarding acceptability, over 93% (29/31 [94%]) of participants in the Txt2Prevent group reported they agreed or strongly agreed that they were satisfied with the program. About 74% (23/31) agreed or strongly agreed that it helped them manage their condition. When asked in semistructured interviews, 17/18 participants said they would recommend the program to other patients with cardiac issues. The participant who said he would not recommend the program stated that his recommendation would depend on whether the person took the time to read the SMS text messages. All but 2 interview participants reported reading every SMS text message, and these reported that while they did read most of the SMS text messages, it was possible they did not read all of them. All interviewed participants said they would be willing to use SMS text messaging again for health purposes.

Table 7. Study protocol feasibility measures.[a]

Feasibility measure	Descriptive assessment
Recruitment	
Months of recruitment	17
Number participants randomized per month, mean (range)	4.5 (0-15)
Ineligible patients, n/N (%)	223/400 (55.8)
Eligible patients who declined to participate, n/N (%)	96/172 (55.8)
Randomization	
Days from discharge to randomization, mean (SD)	2.1 (2.1)
Participants randomized within 7 days of discharge, n/N (%)	74/76 (97)
Follow-up	
Completed packaged follow-up questionnaires, n/N (%)	69/76 (91)
Packaged follow-up questionnaires done by postal mail, n/N (%)	60/69 (87)
Days after discharge to complete packaged follow-up questionnaires, mean (SD)	73 (17)
Completed packaged follow-up questionnaires done within 6 weeks of the 60-day study period, n/N (%)	64/69 (93)
Completed health care resource use follow-up questionnaires, n/N (%)	69/76 (91)
Health care resource use questionnaires by phone, n/N (%)	65/69 (94)
Days after discharge to complete health care resource use questionnaire, mean (SD)	69 (14)
Completed health care resource use questionnaires done within 6 weeks of the 60-day study period, n/N (%)	67/69 (97)
Participants who completed all sets of follow-up questionnaires, n/N (%)	67/76 (88)
Participants who completed no follow-up questionnaires, n/N (%)	5/76 (7)
Questions completed on received questionnaires, n/N (%)	4613/4624 (99.8)

[a]Packaged follow-up questionnaires included Health Education Impact Questionnaire, Cardiac Self-Efficacy Scale, Morisky Medication Adherence Scale, EQ-5D-5L, and EQ-5D-5L Visual Analog Scale.

Discussion

Principal Findings

Our pilot study assessed the impact, feasibility, and acceptability of a 60-day SMS text messaging program in supporting patients with ACS following hospital discharge. In exploratory adjusted analyses, we did not find statistically significant differences on follow-up scores (controlling for baseline scores where applicable) between the Txt2Prevent group and the usual care group in their self-management domains, health-related quality of life, medication adherence, health care resource use, and self-efficacy in adjusted models, except for the "Total plus" domain, which was impacted by an influential outlier. The study protocol was generally feasible, as seen by high adherence to the study protocol targets for randomization time frames and questionnaire completion rates, although recruitment took much longer than estimated. In terms of acceptability, participants reported they generally found the program acceptable and believed it helped them manage their condition.

In our pilot study, we failed to demonstrate the effect of SMS text messaging on our questionnaire outcomes, including the heiQ, CSE, and medication adherence. Previously, 2 interventions using apps reported improvements in heiQ domains over the short term [47,48], although this was not the case in 2 web-based interventions [49,50]. We also did not find improvements in self-efficacy scores, which is in contrast to a study that used SMS text messages and phone calls for patients with CVD [51]. In addition, several previous SMS text messaging studies targeting CVD medication adherence had a positive effect [52]. Other studies assessing SMS text messaging in a CVD population have measured specific risk factors, such as blood pressure and cholesterol. Chow et al [24] reported positive effects on low-density lipoprotein–cholesterol, systolic blood pressure, body mass index, physical activity, and smoking in an SMS text messaging program. However, not all studies have reported positive effects [22]. Zheng et al [53], who used a framework similar to Chow et al's [24], found greater levels of physical activity at 6 months in the intervention group compared with the control group, but did not observe statistically significant (P=.22) effects on blood pressure. Another study investigating the effect of weekly SMS text messages or emails on the primary prevention of CVD risk factors found no improvements after 1 year [54]. Therefore, while many previous studies have reported positive effects, the results are not consistent.

Differences between our findings and others may be due to our intervention's design. Guided by an advisory committee that included clinicians (a cardiac nurse, a family physician, a community pharmacist, and 2 cardiologists), researchers, and 2 patient partners with lived experiences of CVD, the SMS text messages were education based and included prompts that

aligned with hospital messaging and current guidelines [55]. The messages were revised according to feedback from patient focus groups. We wanted to test whether a simple program design (ie, one-way delivery, prespecified order of SMS text messages) was effective before considering more complicated interventions. SMS text messages were designed to broadly apply to patients and did not include tailoring, such as personalization, feedback, or content matching. Incorporating tailored messages could be beneficial as systematic reviews identified that mobile health studies with positive effects often used tailoring [56,57]. In addition, by having a multifactorial focus, we may not have covered topics frequently enough to instigate change. For example, our medication adherence results contrast previous studies reporting positive effects [22], but only 7 of our 48 SMS text messages covered medication-related topics (see SMS text messages sent on day 1, 7, 9, 17, 18, 22, and 24 in Multimedia Appendix 1). The messages were also 1 way only, in part because this required fewer resources to implement. This may be a limitation as some meta-analyses and reviews have reported 2-way messages were more effective, although this is not consistent [57-59]. Chow et al [60] incorporated behavior change techniques such as prompting consequences and self-monitoring of behavior. It is unclear which behavior change techniques are effective, but a future study could consider including more behavior change techniques [61,62].

Regarding the feasibility of the study protocol, we required 17 months to recruit 76 participants instead of the anticipated 6 months. Six months was estimated because there were approximately 750 ACS discharges in the previous year, and a feasibility survey showed 50% (14/28) of patients owned a mobile phone. We assumed 40.0% (300/750) of patients would be eligible and of those 50.0% (150/300) would agree to participate. However, we missed approaching many patients due to restrictions imposed by our ethics board. The research assistant had to obtain bed numbers of patients with ACS from the clinical nurse leader. They then asked the bedside nurse to confirm with the patient if they could explain the study. This required forming strong relationships with clinical staff. Evening recruitment visits also helped as patients were often discharged shortly after returning to the ward from the catheterization laboratory. Ultimately, 43.0% (172/400) of approached patients were eligible and 44.2% (76/172) of eligible patients agreed to participate. More patients declined to participate in our study compared with the 10%-30% refusal rates reported by other CVD SMS text messaging studies that recruited from hospitals or outpatient clinics [23,63]. Many patients refused when initially approached by clinical staff, so we could not document their reasons. Having brief, standardized wording to provide clinical staff could increase uptake. Although the focus groups and feasibility survey indicated patients were interested in this type of program, the hospital environment could have created challenges as patients may be overwhelmed, unsure of their postdischarge needs, or be wary of committing to a research project [64]. However, recruiting outside the hospital would contradict with the time-sensitive nature of the program. The randomization process worked well, although delays happened when questionnaires could not be completed before discharge. Our follow-up rates were slightly lower than previous CVD SMS text messaging studies, which were often between 97% and 91% [53,63,65,66]. Having 2 different questionnaire formats (telephone for 1 questionnaire and mail for the remaining questionnaires) likely caused some of the partial completions.

Limitations

Our study has several limitations to consider. As this was a pilot study, we did not determine our sample size based on power calculations and were likely underpowered to detect clinically important differences. We chose the heiQ as it covers potential proximal and intermediate outcomes of self-management programs [32]. Self-management is important as it is linked to improved health behaviors and reduced costs and health care visits [67-69]. Previously, a 10-week proof-of-concept study (n=35) evaluating a peer-support app reported improvements in heiQ domains, indicating changes are possible in small samples over the short term [48]. Other SMS text messaging studies frequently use clinical measures, such as blood pressure, making it difficult to compare our results directly. In addition, as our measures were self-reported, there may have been biases (eg, social desirability bias or recall bias); however, we primarily used validated questionnaires and confirmed self-reported hospital visits with hospital records. For some measures, a clinically meaningful change has not yet been determined, so we calculated Cohen *d* effect sizes for better comparison [70]. Participants in the intervention group may have been impacted in ways not captured by the questionnaires or had a different perspective at follow-up (eg, been more aware they were not meeting recommendations). While we cannot know this, it is possible as participants in the interviews and acceptability survey provided positive feedback about the program.

Conclusions

In our exploratory analyses, we did not demonstrate any positive effects of the SMS text messaging intervention in terms of self-management, medication adherence, health-related quality of life, cardiac self-efficacy, or health care resource use. The Txt2Prevent program had an intentionally simple design and was acceptable to participants, but design changes may be needed before proceeding to a larger study. The study protocol was feasible to implement, although improvements to the recruitment process are likely required. Future work should investigate the effect of tailoring, multifactor versus single-factor interventions, 2-way versus 1-way SMS text messaging, and the effectiveness of behavior change techniques.

Acknowledgments

The study received funding from the Canadian Institutes of Health Research (CIHR) through a Catalyst Grant for eHealth Innovations (application number 316822). ER was supported to do this work by a Canada Graduate Scholarships–Master's Program from CIHR. BS is supported by a Michael Smith Foundation for Health Research (MSFHR) Scholar Award. MM is

supported by an MSFHR Scholar Award and a CIHR Embedded Clinician Researcher Salary Award. HV is supported by funding from CIHR and the Ontario Ministry of Health and Long-Term Care's Health System Research Fund. SL holds the Pfizer/Heart and Stroke Foundation Chair in Cardiovascular Research Prevention at St. Paul's Hospital, Vancouver, Canada. We acknowledge Mr Daniel Sheinin who was involved in the software development, Ms Nisa Onsel and Ms Joanna Mendell for randomizing study participants, Dr Terry Lee for his statistical advice, the EuroQol Group for their support with the online questionnaire (EQ-5D-5L), Ms Anosha Afaq for her assistance with data entry, and members of the clinical advisory committee, including Ms Alyson Hagan-Johnson (patient partner) and Mr Leon Jung (community pharmacist). We also thank the hospital staff who assisted with recruitment as well as our study participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Txt2Prevent intervention text messages.
[DOCX File , 20 KB - mhealth_v9i5e24530_app1.docx]

Multimedia Appendix 2
Additional analyses - unadjusted Health Education Impact Questionnaire (heiQ), EQ-5D-5L, EQ-5D-5L Visual Analog Scale (EQ VAS), Cardiac Self-Efficacy (CSE), and medication adherence results at 60-days (controlling for baseline scores).
[DOCX File , 23 KB - mhealth_v9i5e24530_app2.docx]

Multimedia Appendix 3
Additional analyses - 60-day follow-up medication prescriptions, by group.
[DOCX File , 14 KB - mhealth_v9i5e24530_app3.docx]

Multimedia Appendix 4
Additional analyses - physician visits and cardiac rehabilitation enrolment within 60-days, by group.
[DOCX File , 14 KB - mhealth_v9i5e24530_app4.docx]

Multimedia Appendix 5
CONSORT E-HEALTH checklist (v1.6.1).
[PDF File (Adobe PDF File), 1738 KB - mhealth_v9i5e24530_app5.pdf]

References

1. Canadian Institute for Health Information. A Snapshot of Health Care in Canada as Demonstrated by Top 10 Lists, 2011. 2012. URL: http://publications.gc.ca/collections/collection_2012/icis-cihi/H115-54-2011-eng.pdf [accessed 2021-04-08]
2. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation 2016 Jan 26;133(4):e38-360. [doi: 10.1161/CIR.0000000000000350] [Medline: 26673558]
3. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA 2013 Jan 23;309(4):355-363 [FREE Full text] [doi: 10.1001/jama.2012.216476] [Medline: 23340637]
4. Southern DA, Ngo J, Martin B, Galbraith PD, Knudtson ML, Ghali WA, et al. Characterizing types of readmission after acute coronary syndrome hospitalization: implications for quality reporting. J Am Heart Assoc 2014 Sep 18;3(5):e001046 [FREE Full text] [doi: 10.1161/JAHA.114.001046] [Medline: 25237046]
5. Chow CK, Jolly S, Rao-Melacini P, Fox KAA, Anand SS, Yusuf S. Association of diet, exercise, and smoking modification with risk of early cardiovascular events after acute coronary syndromes. Circulation 2010 Feb 16;121(6):750-758 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.891523] [Medline: 20124123]
6. Canadian Institute for Health Information. Physician Follow-up after Hospital Discharge: Progress in Meeting Best Practices. 2015. URL: https://secure.cihi.ca/free_products/Physician-Follow-Up-Study-mar2015_EN.pdf [accessed 2021-04-09]
7. Tung Y, Chang G, Chang H, Yu T. Relationship between Early Physician Follow-Up and 30-Day Readmission after Acute Myocardial Infarction and Heart Failure. PLoS One 2017;12(1):e0170061 [FREE Full text] [doi: 10.1371/journal.pone.0170061] [Medline: 28129332]

8. Faridi KF, Peterson ED, McCoy LA, Thomas L, Enriquez J, Wang TY. Timing of First Postdischarge Follow-up and Medication Adherence After Acute Myocardial Infarction. JAMA Cardiol 2016 May 01;1(2):147-155. [doi: 10.1001/jamacardio.2016.0001] [Medline: 27437885]
9. Daugherty SL, Ho PM, Spertus JA, Jones PG, Bach RG, Krumholz HM, et al. Association of early follow-up after acute myocardial infarction with higher rates of medication use. Arch Intern Med 2008 Mar 10;168(5):485-91; discussion 492. [doi: 10.1001/archinte.168.5.485] [Medline: 18332293]
10. Hickson RP, Robinson JG, Annis IE, Killeya-Jones LA, Korhonen MJ, Cole AL, et al. Changes in Statin Adherence Following an Acute Myocardial Infarction Among Older Adults: Patient Predictors and the Association With Follow-Up With Primary Care Providers and/or Cardiologists. J Am Heart Assoc 2017 Oct 19;6(10):e007106 [FREE Full text] [doi: 10.1161/JAHA.117.007106] [Medline: 29051213]
11. Jackevicius CA, Li P, Tu JV. Prevalence, predictors, and outcomes of primary nonadherence after acute myocardial infarction. Circulation 2008 Feb 26;117(8):1028-1036 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.706820] [Medline: 18299512]
12. Hanssen TA, Nordrehaug JE, Hanestad BR. A qualitative study of the information needs of acute myocardial infarction patients, and their preferences for follow-up contact after discharge. Eur J Cardiovasc Nurs 2005 Mar;4(1):37-44. [doi: 10.1016/j.ejcnurse.2004.11.001] [Medline: 15718191]
13. Roebuck A, Furze G, Thompson DR. Health-related quality of life after myocardial infarction: an interview study. J Adv Nurs 2001 Jun;34(6):787-794. [doi: 10.1046/j.1365-2648.2001.01809.x] [Medline: 11422549]
14. Feng L, Li L, Liu W, Yang J, Wang Q, Shi L, et al. Prevalence of depression in myocardial infarction: A PRISMA-compliant meta-analysis. Medicine (Baltimore) 2019 Feb;98(8):e14596 [FREE Full text] [doi: 10.1097/MD.0000000000014596] [Medline: 30813183]
15. Ziebland S, Locock L, Fitzpatrick R. Appendix 4, Myocardial infarction patients? perspectives of care: a secondary analysis of qualitative interviews. Informing the Development of NICE (National Institute for Health and Care Excellence) Quality Standards through Secondary Analysis of Qualitative Narrative Interviews on Patients' Experiences. 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK263661/ [accessed 2021-04-09]
16. Ko DT, Newman AM, Alter DA, Austin PC, Chiu M, Cox JL, et al. Secular trends in acute coronary syndrome hospitalization from 1994 to 2005. Can J Cardiol 2010 Mar;26(3):129-134 [FREE Full text] [Medline: 20352132]
17. Spencer FA, Lessard D, Gore JM, Yarzebski J, Goldberg RJ. Declining length of hospital stay for acute myocardial infarction and postdischarge outcomes: a community-wide perspective. Arch Intern Med 2004 Apr 12;164(7):733-740. [doi: 10.1001/archinte.164.7.733] [Medline: 15078642]
18. Jørstad HT, Minneboo M, Helmes HJM, Fagel ND, Scholte Op Reimer WJ, Tijssen JGP, et al. Effects of a nurse-coordinated prevention programme on health-related quality of life and depression in patients with an acute coronary syndrome: results from the RESPONSE randomised controlled trial. BMC Cardiovasc Disord 2016 Jul 08;16(1):144 [FREE Full text] [doi: 10.1186/s12872-016-0321-4] [Medline: 27391321]
19. Sinclair AJ, Conroy SP, Davies M, Bayer AJ. Post-discharge home-based support for older cardiac patients: a randomised controlled trial. Age Ageing 2005 Jul;34(4):338-343. [doi: 10.1093/ageing/afi116] [Medline: 15955757]
20. Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States. 2021. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2021-05-10]
21. Pew Research Center. Cell Phone Activities. 2015. URL: http://www.pewinternet.org/chart/cell-phone-activities/ [accessed 2021-04-09]
22. Unal E, Giakoumidakis K, Khan E, Patelarou E. Mobile phone text messaging for improving secondary prevention in cardiovascular diseases: A systematic review. Heart Lung 2018;47(4):351-359. [doi: 10.1016/j.hrtlng.2018.05.009] [Medline: 29803297]
23. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: Results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
24. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients With Coronary Heart Disease: A Randomized Clinical Trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
25. Shariful Islam SM, Farmer AJ, Bobrow K, Maddison R, Whittaker R, Pfaeffli Dale LA, et al. Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart 2019;6(2):e001017 [FREE Full text] [doi: 10.1136/openhrt-2019-001017] [Medline: 31673381]
26. Ross ES, Sakakibara BM, Mackay MH, Whitehurst DG, Singer J, Toma M, et al. The Use of Text Messaging to Improve the Hospital-to-Community Transition in Acute Coronary Syndrome Patients (Txt2Prevent): Intervention Development and Pilot Randomized Controlled Trial Protocol. JMIR Res Protoc 2017 May 23;6(5):e91 [FREE Full text] [doi: 10.2196/resprot.6968] [Medline: 28536088]
27. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

28. Dharmarajan K, Hsieh AF, Kulkarni VT, Lin Z, Ross JS, Horwitz LI, et al. Trajectories of risk after hospitalization for heart failure, acute myocardial infarction, or pneumonia: retrospective cohort study. BMJ 2015 Feb 05;350:h411 [FREE Full text] [Medline: 25656852]
29. Grace SL, Tan Y, Marcus L, Dafoe W, Simpson C, Suskin N, et al. Perceptions of cardiac rehabilitation patients, specialists and rehabilitation programs regarding cardiac rehabilitation wait times. BMC Health Serv Res 2012 Aug 16;12:259 [FREE Full text] [doi: 10.1186/1472-6963-12-259] [Medline: 22897912]
30. Smedegaard L, Numé A, Charlot M, Kragholm K, Gislason G, Hansen PR. Return to Work and Risk of Subsequent Detachment From Employment After Myocardial Infarction: Insights From Danish Nationwide Registries. J Am Heart Assoc 2017 Oct 04;6(10):e006486 [FREE Full text] [doi: 10.1161/JAHA.117.006486] [Medline: 28978528]
31. Jackson D, Daly J, Davidson P, Elliott D, Cameron-Traub E, Wade V, et al. Women recovering from first-time myocardial infarction (MI): a feminist qualitative study. J Adv Nurs 2000 Dec;32(6):1403-1411. [doi: 10.1046/j.1365-2648.2000.01622.x] [Medline: 11136408]
32. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
33. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
34. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. Med Care 2016 Jan;54(1):98-105 [FREE Full text] [doi: 10.1097/MLR.0000000000000447] [Medline: 26492214]
35. EuroQol Research Foundation. EQ-5D-5L user guide. 2019. URL: https://euroqol.org/wp-content/uploads/2021/01/EQ-5D-5LUserguide-08-0421.pdf [accessed 2021-04-09]
36. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013 Sep;22(7):1717-1727 [FREE Full text] [doi: 10.1007/s11136-012-0322-4] [Medline: 23184421]
37. Schweikert B, Hahmann H, Leidl R. Validation of the EuroQol questionnaire in cardiac rehabilitation. Heart 2006 Jan;92(1):62-67 [FREE Full text] [doi: 10.1136/hrt.2004.052787] [Medline: 15797936]
38. Sullivan MD, LaCroix AZ, Russo J, Katon WJ. Self-efficacy and self-reported functional status in coronary heart disease: a six-month prospective study. Psychosom Med 1998;60(4):473-478. [Medline: 9710293]
39. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008 May;10(5):348-354 [FREE Full text] [Medline: 18453793]
40. Krousel-Wood M, Islam T, Webber LS, Re RN, Morisky DE, Muntner P. New medication adherence scale versus pharmacy fill rates in seniors with hypertension. Am J Manag Care 2009 Jan;15(1):59-66 [FREE Full text] [Medline: 19146365]
41. Morisky DE, DiMatteo MR. Improving the measurement of self-reported medication nonadherence: response to authors. J Clin Epidemiol 2011 Mar;64(3):255-7; discussion 258 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.09.002] [Medline: 21144706]
42. Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Ganiats TG, Holmes DR, ACC/AHA Task Force Members. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Dec 23;130(25):e344-e426. [doi: 10.1161/CIR.0000000000000134] [Medline: 25249585]
43. Lauck S, Johnson JL, Ratner PA. Self-care behaviour and factors associated with patient outcomes following same-day discharge percutaneous coronary intervention. Eur J Cardiovasc Nurs 2009 Sep;8(3):190-199. [doi: 10.1016/j.ejcnurse.2008.12.002] [Medline: 19162553]
44. Morris SB. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organizational Research Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
45. McNutt L, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003 May 15;157(10):940-943. [Medline: 12746247]
46. Statistics Canada. Census Dictionary: Census Year, 2011. 2015. URL: http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo009-eng.cfm [accessed 2021-04-09]
47. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A Low-Intensity Mobile Health Intervention With and Without Health Counseling for Persons With Type 2 Diabetes, Part 1: Baseline and Short-Term Results From a Randomized Controlled Trial in the Norwegian Part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
48. Sakakibara BM, Ross E, Arthur G, Brown-Ganzert L, Petrin S, Sedlak T, et al. Using Mobile-Health to Connect Women with Cardiovascular Disease and Improve Self-Management. Telemed J E Health 2017 Mar;23(3):233-239. [doi: 10.1089/tmj.2016.0133] [Medline: 27623231]

49. Ammerlaan J, van Os-Medendorp H, de Boer-Nijhof N, Scholtus L, Kruize AA, van Pelt P, et al. Short term effectiveness and experiences of a peer guided web-based self-management intervention for young adults with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2017 Oct 13;15(1):75 [FREE Full text] [doi: 10.1186/s12969-017-0201-1] [Medline: 29029616]
50. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
51. Boroumand S, Moeini M. The effect of a text message and telephone follow-up program on cardiac self-efficacy of patients with coronary artery disease: A randomized controlled trial. Iran J Nurs Midwifery Res 2016;21(2):171-176 [FREE Full text] [doi: 10.4103/1735-9066.178243] [Medline: 27095991]
52. 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. [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
53. Zheng X, Spatz ES, Bai X, Huo X, Ding Q, Horak P, et al. Effect of Text Messaging on Risk Factor Management in Patients With Coronary Heart Disease: The CHAT Randomized Clinical Trial. Circ Cardiovasc Qual Outcomes 2019 Apr;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
54. Anand SS, Samaan Z, Middleton C, Irvine J, Desai D, Schulze KM, South Asian Heart Risk Assessment Investigators. A Digital Health Intervention to Lower Cardiovascular Risk: A Randomized Clinical Trial. JAMA Cardiol 2016 Aug 01;1(5):601-606. [doi: 10.1001/jamacardio.2016.1035] [Medline: 27438754]
55. Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Ganiats TG, Holmes DR, et al. 2014 AHA/ACC Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014 Dec 23;64(24):e139-e228 [FREE Full text] [doi: 10.1016/j.jacc.2014.09.017] [Medline: 25260718]
56. Park LG, Howie-Esquivel J, Dracup K. A quantitative systematic review of the efficacy of mobile phone interventions to improve medication adherence. J Adv Nurs 2014 Sep;70(9):1932-1953. [doi: 10.1111/jan.12400] [Medline: 24689978]
57. Head KJ, Noar SM, Iannarino NT, Grant HN. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
58. 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]
59. 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]
60. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
61. 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]
62. Winter SJ, Sheats JL, King AC. The Use of Behavior Change Techniques and Theory in Technologies for Cardiovascular Disease Prevention and Treatment in Adults: A Comprehensive Review. Prog Cardiovasc Dis 2016;58(6):605-612. [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]
63. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
64. Kandola DK, Banner D, Araki Y, Bates J, Hadi H, Lear SA. The Participant Recruitment Outcomes (PRO) study: Exploring contemporary perspectives of telehealth trial non-participation through insights from patients, clinicians, study investigators, and study staff. Contemp Clin Trials Commun 2018 Sep;11:75-82 [FREE Full text] [doi: 10.1016/j.conctc.2018.05.005] [Medline: 29998203]
65. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet J, Alessi M, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]
66. Khonsari S, Subramanian P, Chinna K, Latif LA, Ling LW, Gholami O. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. Eur J Cardiovasc Nurs 2015 Apr;14(2):170-179. [doi: 10.1177/1474515114521910] [Medline: 24491349]
67. Guo P, Harris R. The effectiveness and experience of self-management following acute coronary syndrome: A review of the literature. Int J Nurs Stud 2016 Sep;61:29-51. [doi: 10.1016/j.ijnurstu.2016.05.008] [Medline: 27267181]
68. Wheeler JRC, Janz NK, Dodge JA. Can a disease self-management program reduce health care costs? The case of older women with heart disease. Med Care 2003 Jun;41(6):706-715. [doi: 10.1097/01.MLR.0000065128.72148.D7] [Medline: 12773836]

69. Panagioti M, Richardson G, Small N, Murray E, Rogers A, Kennedy A, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014;14:356 [FREE Full text] [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
70. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T, Clinical Significance Consensus Meeting Group. Estimating clinically significant differences in quality of life outcomes. Qual Life Res 2005 Mar;14(2):285-295. [doi: 10.1007/s11136-004-0705-2] [Medline: 15892420]

Abbreviations

ACS: acute coronary syndrome
ANCOVA: analysis of covariance
CSE: Cardiac Self-Efficacy scale
CVD: cardiovascular disease
EQ VAS: EQ-5D-5L Visual Analog Scale
MMAS: Morisky Medication Adherence Scale

Edited by L Buis; submitted 23.09.20; peer-reviewed by M Karunanithi, HL Tam; comments to author 02.12.20; revised version received 27.01.21; accepted 11.03.21; published 14.05.21.

Please cite as:
Ross ES, Sakakibara BM, Mackay MH, Whitehurst DGT, Singer J, Toma M, Corbett KK, Van Spall HGC, Rutherford K, Gheorghiu B, Code J, Lear SA
The Use of SMS Text Messaging to Improve the Hospital-to-Community Transition in Patients With Acute Coronary Syndrome (Txt2Prevent): Results From a Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e24530
URL: https://mhealth.jmir.org/2021/5/e24530
doi:10.2196/24530
PMID:33988519

Original Paper

Developing Messaging Content for a Physical Activity Smartphone App Tailored to Low-Income Patients: User-Centered Design and Crowdsourcing Approach

Laura Elizabeth Pathak[1], BA, MSW; Adrian Aguilera[1,2,3], PhD; Joseph Jay Williams[4], PhD; Courtney Rees Lyles[2,5], PhD; Rosa Hernandez-Ramos[1], BA; Jose Miramontes[2,5], BA; Anupama Gunshekar Cemballi[2,5], MA; Caroline Astrid Figueroa[1], MD, PhD

[1]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[2]Center for Vulnerable Populations, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States

[4]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[5]Division of General Internal Medicine, Zuckerberg San Francisco General Hospital, San Francisco, CA, United States

Corresponding Author:
Adrian Aguilera, PhD
School of Social Welfare
University of California, Berkeley
120 Haviland Hall, MC 7400
Berkeley, CA, 94720
United States
Phone: 1 510 642 8564
Email: aguila@berkeley.edu

Abstract

Background: Text messaging interventions can be an effective and efficient way to improve health behavioral changes. However, most texting interventions are neither tested nor designed with diverse end users, which could reduce their impact, and there is limited evidence regarding the optimal design methodology of health text messages tailored to low-income, low–health literacy populations and non-English speakers.

Objective: This study aims to combine participant feedback, crowdsourced data, and researcher expertise to develop motivational text messages in English and Spanish that will be used in a smartphone app–based texting intervention that seeks to encourage physical activity in low-income minority patients with diabetes diagnoses and depression symptoms.

Methods: The design process consisted of 5 phases and was iterative in nature, given that the findings from each step informed the subsequent steps. First, we designed messages to increase physical activity based on the behavior change theory and knowledge from the available evidence. Second, using user-centered design methods, we refined these messages after a card sorting task and semistructured interviews (N=10) and evaluated their likeability during a usability testing phase of the app prototype (N=8). Third, the messages were tested by English- and Spanish-speaking participants on the Amazon Mechanical Turk (MTurk) crowdsourcing platform (N=134). Participants on MTurk were asked to categorize the messages into overarching theoretical categories based on the capability, opportunity, motivation, and behavior framework. Finally, each coauthor rated the messages for their overall quality from 1 to 5. All messages were written at a sixth-grade or lower reading level and culturally adapted and translated into neutral Spanish by bilingual research staff.

Results: A total of 200 messages were iteratively refined according to the feedback from target users gathered through user-centered design methods, crowdsourced results of a categorization test, and an expert review. User feedback was leveraged to discard unappealing messages and edit the thematic aspects of messages that did not resonate well with the target users. Overall, 54 messages were sorted into the correct theoretical categories at least 50% of the time in the MTurk categorization tasks and were rated 3.5 or higher by the research team members. These were included in the final text message bank, resulting in 18 messages per motivational category.

Conclusions: By using an iterative process of expert opinion, feedback from participants that were reflective of our target study population, crowdsourcing, and feedback from the research team, we were able to acquire valuable inputs for the design of

motivational text messages developed in English and Spanish with a low literacy level to increase physical activity. We describe the design considerations and lessons learned for the text messaging development process and provide a novel, integrative framework for future developers of health text messaging interventions.

(JMIR Mhealth Uhealth 2021;9(5):e21177) doi:10.2196/21177

KEYWORDS

user centered design; mHealth; text messaging; crowdsourcing; mobile phone

Introduction

Background

Depression and diabetes are both highly disabling and often comorbid diseases that disproportionately affect patients with a low-income and ethnic minority status. For instance, ethnic and racial minority patients generally show a higher prevalence of these diseases [1], lower treatment rates [2,3], and worse outcomes [4,5]. There is a need for the design of more effective self-management interventions that can target both these diseases and are affordable, acceptable, and tailored to vulnerable populations [6]. Promoting physical activity is a potentially effective strategy that has positive effects on both mental health, with a particularly strong effect on depression [7,8], and common chronic diseases such as diabetes [9].

Mobile phone SMS, or text messaging, interventions have shown great promise in helping individuals engage in healthy behaviors, including physical activity [10,11] and diabetes self-management [12-14]. Most individuals currently own a mobile phone. Smartphone usage across a wide range of demographic groups is high (around 81%) and is on the rise. For instance, the percentage of Black individuals (80%) and Latino individuals (79%) that own smartphones is now similar to that of White individuals (82%) [15]. Therefore, feedback and motivational text messages might be effective in helping individuals with depression and diabetes increase their levels of physical activity and improve their overall health.

However, there is limited evidence regarding the design process of text messages aimed at promoting behavioral changes [16]. Detailed accounts of complete content development processes, including formative research and pretesting methods, are underreported in the mobile health (mHealth) literature [17]. Among the published health text messaging studies that describe their content design processes, there is great variability in the methods utilized. Researchers have reported the development of messaging based on public health guidelines [18-21], health education curricula [14,22], theoretical models [16,18,20,21,23], findings from quantitative surveys or focus groups [16,18,19,21,23,24], evaluation by members of the research team or outside experts [16,18,20,22], and pilot testing within a subset of the target population [19,20,25]. Thus, this study aims to integrate various methods such as participant feedback, pilot testing, crowdsourcing, and expert knowledge.

Moreover, very little has been described about the design process of text messages tailored to low-income, low–health literacy populations and non-English speakers. However, the current consensus is that digital interventions should emphasize usability and engagement with content [26] and should be developed with users who are intended to benefit from the intervention [27]. This perspective is fundamental to user-centered design (UCD), a design approach derived from the multidisciplinary field of study of human-computer interaction. UCD entails the active participation of end users in product development to enhance the understanding of user and task requirements as well as the iteration of design and evaluation [28].

In the context of health text messaging development, a UCD approach consists of conducting preliminary research within the target population to inductively identify any barriers that they face and collect feedback about the content as it is being developed [21,29,30]. However, recruiting large groups of participants within patient populations to test the content is challenging. Online crowdsourcing platforms, such as Amazon's Mechanical Turk (MTurk), are increasingly being used as a means of acquiring feedback from a relatively large pool of participants. However, relying solely on MTurk respondents might not be entirely feasible as these participants are not always reflective of typical patient populations. Therefore, specifically for research on the design of text messaging in underserved populations, an argument can be made to combine crowdsourcing with participant feedback.

Objectives

Considering the limited literature available and the lack of consensus on best practices for designing a text messaging intervention, particularly for low-income Spanish-speaking populations, we aimed to develop messaging that was both evidence based and responsive to user feedback. Here, we describe the iterative design process of text messages for use within an adaptive smartphone app for low-income ethnic minority patients. We present a novel framework for health text messaging design, the theory-informed UCD framework, which integrates a UCD approach with crowdsourcing and expertise within the study team. In seeking to advance the body of knowledge of mHealth development methods, we provide recommendations and *lessons learned* from designing text messages in this population.

Methods

Overview

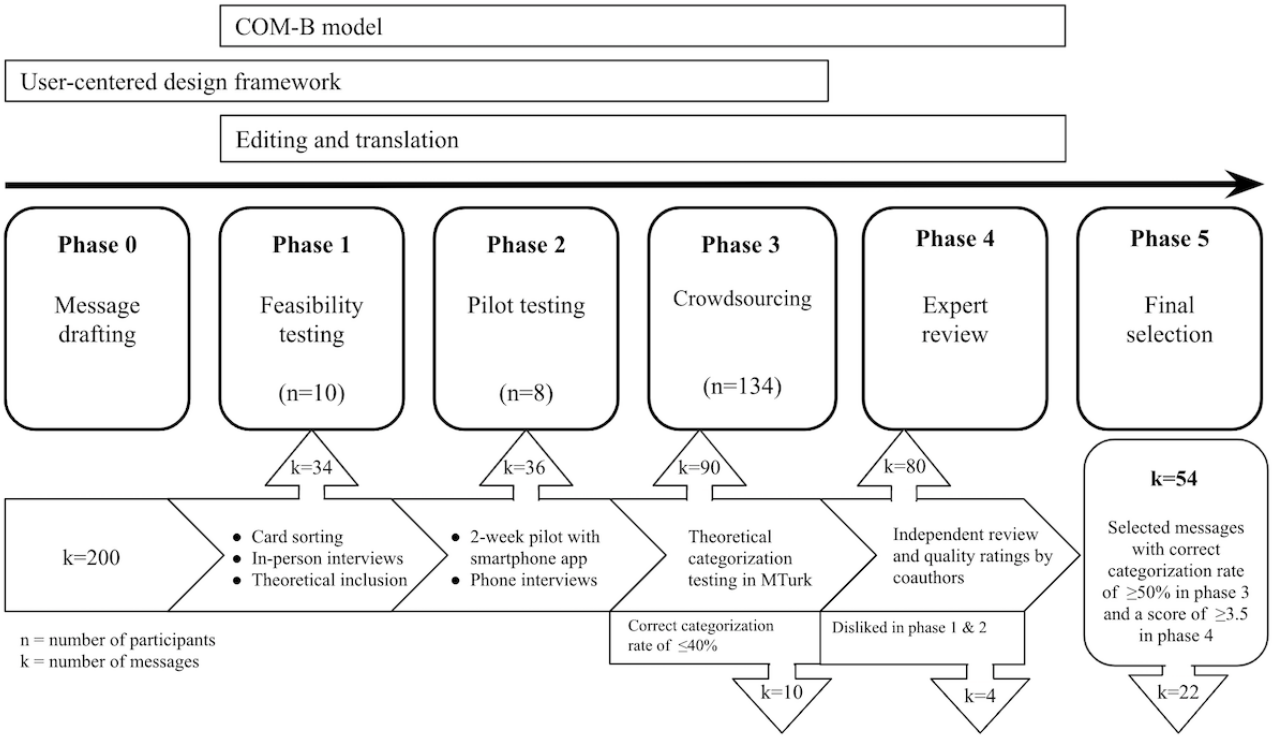
This text messaging design is part of a larger randomized clinical trial, the *Diabetes and Mental Health Adaptive Notification Tracking and Evaluation* (DIAMANTE) trial (NCT 03490253) [31]. This trial examines the effectiveness of a smartphone app that uses a reinforcement learning algorithm to predict the categories of text messages that are most effective in increasing a participant's physical activity, given the participant's

contextual variables, such as previous physical activity, as well as demographic and clinical variables (eg, age, gender, and depression scores) [32]. Although we plan to report the effectiveness results of the trial in future studies, this paper explains the development process of the different categories of motivational text messages for increasing physical activity targeted to individuals with comorbid diabetes and depression. The text messages were developed in 5 phases that incorporated crowdsourced feedback, study team feedback, and UCD techniques (ie, card sorting, interviews, and surveys). The process was iterative in nature as the findings from each step informed the subsequent steps.

Message Development Phases

The messages were created in English and Spanish and tested in 5 phases (Figure 1).

Figure 1. Theory-based, user-centered design framework: phases and methods of the message development process. COM-B: capability, opportunity, motivation, and behavior; MTurk: Amazon Mechanical Turk.



Phase 0: Message Drafting

This phase occurred between August 2017 and June 2018. To ground our intervention in an understanding of the enablers and barriers to behavior change, we developed a library of concise messages (n=200) based on health behavior change constructs from social cognitive theory and evidence-informed guidelines for mHealth interventions found in the literature (eg, *gain and loss* framing, proximal outcomes framing) [33-37].

Phase 1: Feasibility Testing

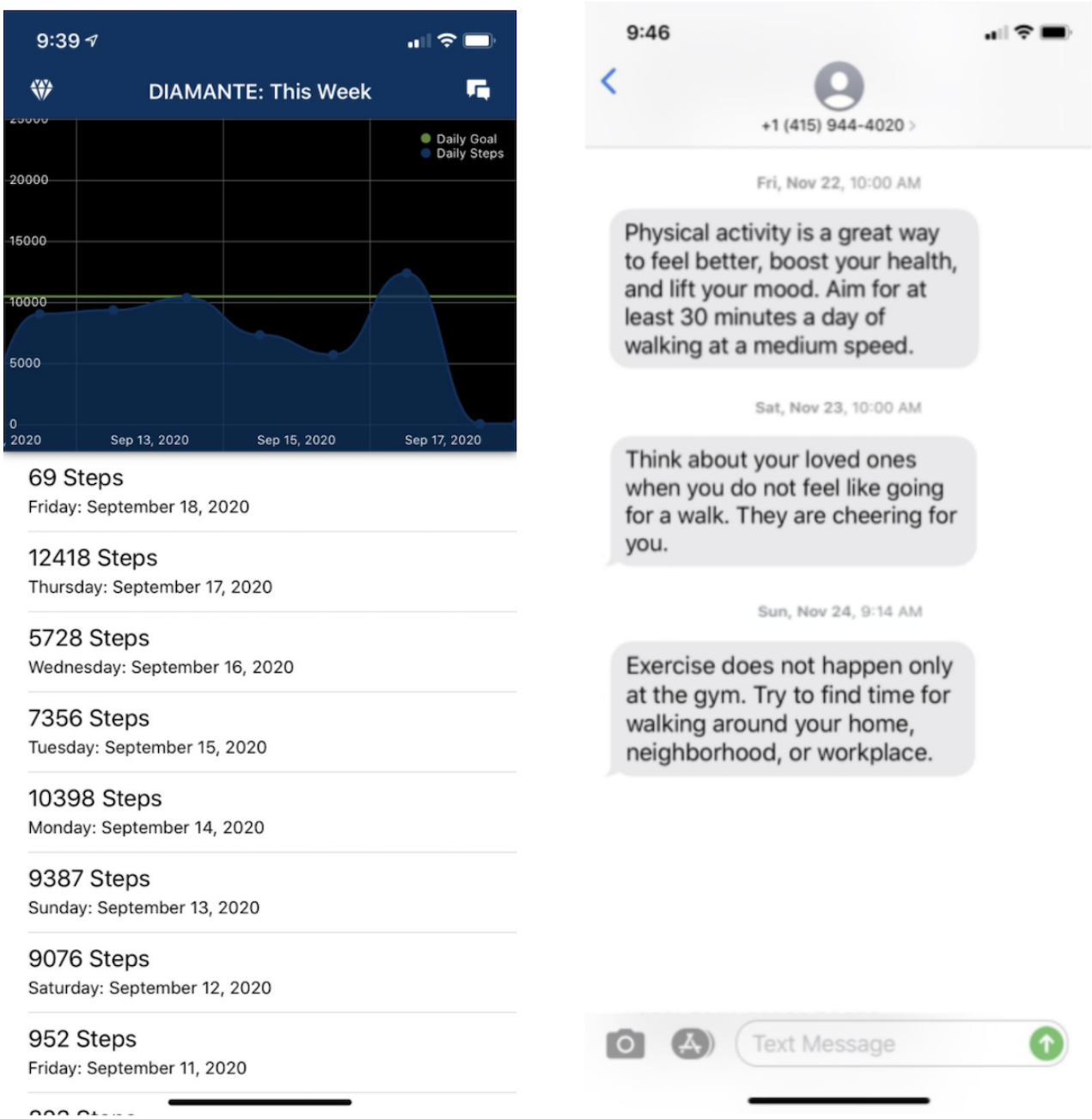
This phase took place between July and December 2018. It consisted of a feasibility study phase with English-speaking (n=5) and Spanish-speaking (n=5) primary care adult patients recruited from a safety-net health care setting. All participants (10/10, 100%) were diagnosed with depression, and 90% (9/10) had diabetes [30,38]. A subset of messages (n=34) was tested using a mixed methods approach that included a card sorting task and individual semistructured interviews. For the card sorting activity, participants were asked to sort text messages in either English or Spanish (depending on their preferred language) into 3 piles: liked messages, disliked messages, and messages that were neither disliked nor liked (neutral). Participants also provided the reason behind the placement of the card in a particular pile. Although card sorting has been previously used to validate expert-developed health text messages among a group of nonexperts (ie, university students and staff) [39], the use of this method in our study entailed modification from the traditional card sorting protocol to adapt it to our population of low-income patients [40]. For example, the subset of messages represented in the cards was written at a sixth-grade reading level. In addition, interviewers modified the administration of the activity to accommodate participants with limited literacy or communication barriers by providing audiovisual cues, including reading the cards aloud and successively probing for feedback after reading each card [40].

The individual semistructured interviews entailed a set of open-ended questions that ultimately informed the thematic content of messages and messaging characteristics [30]. Participants were also asked about their perceived barriers and facilitators of regular physical activity. Upon conclusion of this feasibility phase, all messages (n=200) were further refined to incorporate participant feedback and a cognitive framework for behavior change—the capability, opportunity, motivation, and behavior (COM-B) model—and edited and translated by members of the research team.

Phase 2: Pilot Testing

This phase took place between January and April 2019. It consisted of a 2-week technology acceptance pilot with English-speaking (n=4) and Spanish-speaking (n=4) patients with comorbid diabetes and depression recruited from the same primary care setting as phase 1. Participants were enrolled on a rolling basis. The pilot entailed testing a subset of messages (n=36), which were updated based on findings from phase 1 (ie, patient feedback from the card sorting activity and interviews). Importantly, these messages were tested within the smartphone app prototype developed by authors AA and CRL and Audacious Software for the DIAMANTE trial [32] (Figure 2). A total of 12 messages from each theoretical category of the full message bank were selected for testing to ensure that the subset (n=36) was representative of our theoretical construct categorization. Moreover, because we wanted to maintain some variability in messaging to evaluate the impact of the adaptive learning algorithm, any messages that received negative feedback in the previous phase were still tested in this phase. The study team members monitored the messages received by the participants. In addition, participants were called by a team member every weekday during business hours at a selected time chosen by the participant to ask them about their experiences with the app and to provide qualitative feedback on the messages. Individualized implementation methods have been recommended for enhancing participant buy-in in maintaining commitment to the intervention once enrolled and minimizing dropout rates in mHealth studies [41,42]. Therefore, in addition to eliciting regular and reliable feedback from participants on the messaging content, we used a daily phone call strategy to enhance engagement with the app and minimize dropouts during pilot testing, particularly among Spanish-speaking participants and those with limited health and digital literacy.

Figure 2. Interface of the smartphone app and intervention text messages in a test user's texting interface.



Phase 3: Crowdsourcing

This phase took place between February 2019 and March 2019. A subset of 90 messages (30 per category) composed of the 36 messages tested in phase 2 and 54 chosen at random was surveyed among MTurk participants for motivational construct categorization (Table 1). Although a handful of messages (n=4) tested in phases 1 and 2 received predominantly negative participant feedback, we still included these messages in the MTurk sample for variability to assess the impact of the algorithm. As educational attainment and income are correlated [43], we selected participants with high school education or less to inform an intervention targeting low-income users. MTurk participants were asked to read different messages and determine their category (among the 3 overarching theoretical categories: benefit, self-efficacy, and opportunity cues) to the best of their abilities, based on the examples and descriptions provided. Categorization questions were set up as multiple-choice questions with the 3 categories as possible options. The order of the options was not randomized, but the sequence of the questions was randomized. Correct rates (scores) of the English and Spanish categorization questions were calculated for each participant.

Table 1. Demographic data of Amazon Mechanical Turk participants (N=134).

Demographic category	Value, n (%)
Language	
English	117 (87.3)
Spanish	17 (12.7)
Location	
United States	119 (88.8)
Mexico	4 (3)
South America	9 (6.7)
Spain	2 (1.5)

Phase 4: Expert Evaluation

This phase occurred in April 2019. After discarding messages (n=10) based on poor performance in phase 3, 80 preselected messages were rated for quality by the study team members using a 5-point Likert-type scale ranging from 1 (poor quality) to 5 (high quality). These ratings were independently performed to reduce bias. In addition, each team member was able to see the entire original set of messages (n=200) and all data from phases 1, 2, and 3.

Phase 5: Final Selection

A final bank of messages (n=54) was selected based on team ratings, MTurk scores, and participant feedback. Analysis of the final message selection and final content editing took place from June to August 2019.

Results

Phase 0: Message Drafting

We operationalized the constructs of social cognitive theory (ie, self-efficacy and outcome expectancies) [33] into practical cues and tips for physical activity in a subset of messages. In addition, given past research that shows that exposure to proximal positive exercise outcomes increases intrinsic motivation among individuals who report lower levels of past physical activity [34], we framed most messages in terms of proximal outcomes and underscored the positive health-related feelings associated with exercising [35]. Furthermore, we wrote half of the messages with a *gain* framing, emphasizing the benefits of physical activity, and half with a *loss* framing, emphasizing what is avoided by doing physical activity [36]. Health communication research has shown that *gain and loss* framing has the capacity to both increase engagement with message content and change behavior [37,38]. Finally, we wrote messages that followed *individualistic*, *family*, and *peer* orientations to cue participants to increase their current physical activity levels as well as messages that addressed environmental constraints for physical activity [44].

The messages at this stage (n=200) were in a draft form. We aimed to refine the content after gaining a better understanding of the barriers of and beliefs about physical activity of our target population in the subsequent UCD phases, as we anticipated some discrepancy between participants' identified beliefs and barriers and those we found in the literature. We also sought to identify any silent diabetes- and depression-related barriers for exercise and how well the individualistic, family, and peer orientations resonated with participants.

Phase 1: Feasibility Testing

Phase 1 consisted of implementing the UCD framework of end user (patient) involvement to obtain qualitative feedback that would allow us to make design decisions. Detailed results of this phase have been published previously [30,40].

Card Sorting

Feedback on the existing content consisted of liking messages that allowed for self-reflection, provided concrete ideas for engaging in physical activity, and were highly motivating. In general, participants disliked messages that they perceived as repetitive. Participants also suggested that sticking to highly motivating messages and personalized advice would strengthen the content. Overall, out of the 34 messages evaluated in this phase, 5 (15%) were predominantly liked (by 3 or more participants), 2 (6%) were predominantly considered neutral, and only 1 (3%) message was predominantly disliked.

Qualitative Interviews

Key findings revealed 3 main health-related barriers: physical limitations, chronic pain, and depression. We found that the primary facilitators of physical activity included being able to clearly visualize or understand the health-related benefits of engaging in regular physical activity, self-motivation and encouragement to change from a sedentary to an active lifestyle, and offering advice on how to engage in regular physical activity.

Theoretical Inclusion

On the basis of participant feedback from the card sorting activity and qualitative interviews, we identified the COM-B model [41] as the optimal theoretical framework for message categorization and conceptualization. According to the COM-B model, behavior change is part of an interacting system involving the key ingredients of capability, opportunity, and motivation [45]. This model lays out a mechanism for the development of effective behavioral health interventions: changing one or more of these components to rearrange the system and minimize the risk of default [45]. Although the COM-B model draws from social cognitive theory, which was

the original chosen theoretical base for the messages, we incorporated the COM-B model because it allowed for more optimal linking with our intervention design, as it was a better fit for the opportunity cues to be organized within our message development process.

Accordingly, we reframed our library of draft messages to fit into each of the 3 COM-B model constructs [30]. In targeting the *capability* construct, which is influenced by internal processes that direct behavior, we reconstructed the messages that provided information on the determined benefits of physical activity to highlight physical and social outcome expectations as motivators for physical activity (72/200, 36%). The *opportunity* construct explains the social or environmental factors that allow change to occur. Therefore, we decided to redefine existing messages that addressed environmental constraints for physical activity to emphasize cues for exercise opportunities or suggestions on how to reduce or circumvent these constraints (eg, restructuring routines and planning for social support; 73/200, 36.5%). Finally, for the *motivation* construct, which is demonstrated by the skills and knowledge necessary for change, we used existing messages that targeted self-efficacy (55/200, 27.5%) through tips for building self-management skills and self-beliefs of motivation and confidence to exercise (Table 2).

Table 2. Sample of final selected messages by theoretical construct category.

COM-B[a] construct	Motivational orientation	
	Individual	Social
Capability (benefits of physical activity)	• Physical activity is a great way to feel better, boost your health, and lift your mood. Aim for at least 30 minutes a day of walking at a medium speed. • Managing your body pain can be hard. Exercise can help! • Going for a walk can improve your mood and clear your mind. • It may be painful to start walking, but walking often can help with pain relief.	• When you are active, you have more energy to spend time with your loved ones. • Your loved ones will be proud to see the changes in your health over time. • You are important to your loved ones. Being more active can help you take care of them. • Being active can improve your mood. This can help you spend quality time with your loved ones.
Opportunity (cues for engaging in physical activity)	• Exercise does not happen only at the gym. Try to find time for walking around your home, neighborhood, or workplace. • Try not to sit for more than 30 minutes at a time. You can walk around your home for a few minutes. • Too many things going on? Take a quick walk to destress and take care of your health. • You do not need fancy things to get active. Comfortable shoes and a water bottle should work!	• Ask your loved ones for support to be more active. • Invite a loved one to go on a walk with you. Use this time to catch up! • Change your routine. Go for a walk with loved ones after dinner instead of watching TV. • Think about walking as a way for you to visit new places or parks with your loved ones.
Motivation (self-efficacy for physical activity)	• Be proud of yourself for making small changes. It is not always easy. • You are a strong person. Use that strength to keep trying to reach your walking goal. • Believing in yourself is the first step toward reaching your goal. • Do not feel bad if you cannot walk very far. Start slow. You will get better!	• You can inspire others by being active and staying healthy. • Other people will start to see the changes you are making for your health. • Think about your loved ones when you do not feel like going for a walk. They are cheering for you. • Do not worry if you are not walking as much as other people. Everyone has their own speed.

[a]COM-B: capability, opportunity, motivation, and behavior.

In addition, based on the findings of common constraints for physical activity among our target population, we integrated content that touched on physical symptoms, such as pain, and reframed messages to target the improvement of specific depressive symptoms with the purpose of increasing behavioral activation.

Phase 2: Pilot Testing

Overview

In the initial stage of this pilot, where we evaluated a representative subset of 36 messages (12 for each of the 3 theoretical categories) streamed through the smartphone app prototype, we found that patients tended to not carry their phone or not pay attention to the notification of text messages. However, patients generally started paying more attention to their phones and taking their phones with them more often after a few days into the pilot. In addition, although patients did not always remember receiving text messages when called by the research assistant, all patients reported liking the messages. A total of 75% (6/8) of patients reported feeling indifferent about the timing of the messages.

Furthermore, 38% (3/8) of patients reported not liking the messages in which they were compared with others (eg, other people have the same medical conditions and walked more). Messages about exercising with family members or about the benefits of being healthy to support family members were disliked by 25% (2/8) of participants, who reported not having strong family connections. However, all patients reported that

they felt like they started walking more and that they wanted to continue to receive messages past their 2-week trial.

Following the completion of the pilot and based on the negative appraisal among some patients toward messages alluding to family member support, we removed any references to "family," "peers," or "friends" and replaced them with a blanket "loved ones" phrase. In the same vein, we decided to merge the *family* and *peer* orientations as *social* and therefore reclassify the messages as either *individualistic* or *social*.

Editing

Message editing was performed by the first author. The purpose of editing was the overall quality assurance of all existing messages. The first round of edits focused on identifying and correcting grammar, spelling, and punctuation errors as well as simplifying syntax to a lower reading level. All messages were adapted to a reading level of sixth grade or below (average=3.5). A web-based readability tool [46] was used to measure the reading level of each message using the Flesch-Kincaid grade level formula. Messages with scores higher than 6 were edited until they scored lower than the sixth-grade cutoff. A second round of edits focused on rewording to make the messages more concise to improve the experience of patients reading them. As the number of characters that can fit in a standard SMS is 160, messages with more than 160 characters were edited and reduced to fit this parameter. However, as the research staff had identified patient preferences for shorter messages in the previous phases, most messages were already under 140 characters.

Translation From English Into Spanish

The English text messages were translated into Spanish by the bilingual research staff. The translated messages were also reviewed by a native Spanish speaker fluent in English. In general, all translations were performed following a culturally sensitive communication standard [47]. We adapted syntax, lexical content, and idiosyncratic phrases that have been identified as key cultural aspects of translation in previous text messaging intervention research [47]. This means that the messages were not translated literally; rather, they were adapted and reframed to consider the ethnic diversity and other important cultural characteristics of one of the target populations of the DIAMANTE study—Spanish-speaking immigrants from Latin American countries. As such, neutral Spanish [48] words were used to the extent possible, and formal Spanish was used throughout (ie, using "usted" and its derivations in second person pronouns). A translation key was developed to standardize commonly translated phrases or words (Multimedia Appendix 1). During editing, any changes to English messages were simultaneously applied to the translated Spanish messages. Similar measures to reduce the literacy level of English content were applied to the Spanish content.

Phase 3: Crowdsourcing

In this phase, we submitted a representative sample of messages (n=90, 30 for each theoretical category) to the MTurk participants to test and validate the theoretical categorization. MTurk testing enabled us to ensure that the messages had been tagged in the right motivational categories and, therefore, refine the content and select the best ones. Messages were placed in the right category at an average of 58% (SD 0.21%) of the time. The MTurk respondents also rated messages for their overall quality from 1 to 5. The highest-rated messages were those in the *benefit* category. A total of 70% (63/90) of the messages were put in the right category at least 50% of the time. Furthermore, 11% (10/90) of the messages yielded a mean percent correct categorization rate of 40% or less. These messages were discarded from subsequent evaluations because of their thematic ambiguity. Participants completed the survey in an average of 6.5 minutes (Table 3).

Table 3. Amazon Mechanical Turk classification results of messages by theoretical category (N=30).

Theoretical category	Correct classification, mean n (%)
Benefit	21 (70)
Self-belief	16 (53)
Opportunity	15 (50)

A challenge with using MTurk was to recruit a larger sample of Spanish-speaking workers to test the theoretical categorization of the messages. Otherwise, the task requirements of our categorization tests were easily implemented at the interface of the platform. However, as we found that the qualitative feedback about the message content from the testers was either incomplete or not helpful (ie, not reflective of the needs of our target population), we decided to discontinue this task after the second round of testing.

Phase 4: Expert Evaluation

To assess the content adequacy of the messages, all coauthors independently rated the high-performing messages of phase 3 (n=80) for subjective quality using a scale of 1 to 5. The coauthors gave the messages an average rating of 3.6 (SD 0.59) and reached a quality rating consensus for a given message only 3% (2/80) of the time. A total of 60% (48/80) of the messages received a rating of 3.5 and higher. Quantitative ratings and qualitative feedback were collected from each coauthor and analyzed in aggregate.

Phase 5: Final Selection

Feedback from all 4 phases was combined to select the best messages from the preselected subset from phase 4 (n=80). Messages that compared patients with others with the same medical conditions, which were predominantly disliked by participants in phases 1 and 2, were discarded (4/80, 5%). Of the remaining messages, those that were both sorted into the correct theoretical categories at least 50% of the time during phase 3 and rated at least 3.5 or higher in quality during phase 4 were considered as the clearest and highest impact content to include in the final text message bank. This combined feedback

resulted in a final bank of 54 messages, with 18 messages per motivational category.

Discussion

Principal Findings

Using a theory-based UCD framework, which consists of an iterative development process blending UCD methods with crowdsourcing and expert input, we produced a library of 54 motivational messages for a physical activity SMS intervention for low-income minority patients with comorbid diabetes and depression over a 2-year period. The development process comprised 5 phases: feasibility testing, pilot testing, crowdsourcing, expert evaluation, and final selection.

Lessons Learned

Patient feedback gathered in the UCD phases (1 and 2) was overwhelmingly positive: most patients reported no preference for the timing of message delivery, were receptive to receiving motivational text messages, and reported wanting to continue receiving messages upon conclusion of the pilot testing. However, user feedback is prone to response bias, and it is possible that such positive responses were because of patients not wanting to disappoint the study investigators. Findings from the UCD phases were especially helpful for discarding unappealing messages and editing thematic aspects of messages that did not resonate well with patients. For example, after phase 2, we rephrased any references to peer or familial relations to a broader social denomination (*loved ones*) based on the finding that participants who reported not having a family disliked messages that alluded to family support. The number of content changes we incorporated over 2 rounds of UCD testing underscores the importance of taking a user-centered approach to the development of a text messaging intervention. Results from the MTurk testing phase 3 showed that this platform is a relatively inexpensive, accessible, and rapid source of data for the validation of message classification. Thus, we learned that crowdsourcing methods are valuable for certain aspects of the design process in which a larger group of participants is desired or when seeking specific feedback that is not about engagement with content, such as for determining the correct thematic categorization of messages. Therefore, it is valuable to leverage crowdsourcing platforms as a more accessible and inexpensive source of feedback. Finally, we found that expert input and review (phase 4) rendered cohesiveness and reliability to the design process, given that this stage consisted of making independent ratings of the final set of messages to achieve consensus on the best ones. This was particularly important when it came to translating theory and evidence into succinct and engaging yet actionable and motivational messaging.

Strengths and Limitations

The strengths of this study include the multistaged and evidence-based development of cogent messaging that is grounded in the available scientific best practices, health behavior change and user engagement theories, and feedback from target users. This process allowed us to design content to match and adapt to the relevant needs of our patient population. As such, we wrote all messages at a sixth-grade or lower reading level, took cultural and linguistic appropriateness into account during editing and translation, and tested for acceptability and clarity among target patients. However, this study has some limitations. For instance, the sample sizes for the UCD phases were small and from one geographic location. However, the purpose of qualitative user research is not to produce findings that can be generalized to other populations but rather to generate data about attitudes or behaviors based on direct observation. Indeed, the UCD phases in our study were used to assess the acceptability and usability of a set of motivational text messages for use in a future physical activity intervention study tailored to a diverse clinical sample of low-income patients with diabetes and depression. A second limitation of our study is that although we did not collect demographic information beyond language and the country of residence, MTurk respondents are not likely to be representative of our target population. Thus, despite the aforementioned benefits of sampling from this crowdsourcing platform, findings from phase 4 were interpreted with caution, only used for theoretical category validation, and not used to inform subsequent content changes. A third limitation is that MTurk respondents did not receive the full set of draft messages (n=200) in the theoretical category validation task, which means that we were not able to collect categorization ratings for every message. However, we had raters assess only a subset of messages (n=90) as a measure to reduce the overall burden of the task and not overwhelm the respondents. We also noted messages with motivational construct ambiguity that were not tested in MTurk during the internal expert review phase and later discarded them from the final selection. A fourth limitation is that the MTurk categorization responses are subject to method bias as a function of low motivation; that is, participants may have not been willing or able to expend the required amount of cognitive effort and thus were less thorough in making accurate response selections. Online recruitment methods such as MTurk have been criticized as prone to yield samples full of low effort respondents [49]. However, we operationalized measures to reduce such bias in this content validation phase, including providing clear and precise definitions of the 3 motivational categories and providing a *trial* run during the onboarding stage for MTurk participants.

Future Directions

This paper describes in detail the process for designing and validating the content of an mHealth intervention that utilizes an adaptive learning algorithm to deliver optimized and tailored motivational text messages that promote physical activity. Previous studies have utilized similar methods to inform health text messaging development, such as message drafting based on behavior change theory [16,23] and public health guidelines [18], survey responses of participants' preferences [23], focus groups and expert evaluation [18], and end user ratings of content understanding and appeal [16]. However, the process described herein is unique in that it integrates various design methods previously reported in literature and incorporates crowdsourcing via MTurk as an alternative content pretesting method. This is important because beyond a focus on usability, engagement, and implementation, alternative methods for intervention development in general and text messaging content in particular are a largely untapped area of inquiry in the field

of mHealth. This paper thus helps advance mHealth research addressing intervention design processes, including content pretesting methods, which are not commonly detailed in the literature [17]. However, there continues to be a need for research on text messaging intervention design methodology, especially that which explores how to best integrate behavioral health theory in messaging content. There is also a need to continue evaluating new empirical models of content testing, including those that leverage crowdsourcing and online recruitment methods such as Amazon's MTurk and Facebook.

Conclusions

We present a novel, integrative framework for combining UCD, expert input, and crowdsourcing to determine, create, and improve the content of text messages of a smartphone app to increase physical activity in low-income, English- and Spanish-speaking patients with depression and diabetes. This provides a research-based design approach for future developers of health text messaging interventions. This framework can potentially be extended to mHealth programs targeting other vulnerable populations and health behaviors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Translation key.
[DOCX File , 7 KB - mhealth_v9i5e21177_app1.docx]

References

1. Thomas J, Jones G, Scarinci I, Brantley P. A descriptive and comparative study of the prevalence of depressive and anxiety disorders in low-income adults with type 2 diabetes and other chronic illnesses. Diabetes Care 2003 Aug;26(8):2311-2317. [Medline: 12882854]
2. Alegría M, Chatterji P, Wells K, Cao Z, Chen C, Takeuchi D, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. Psychiatr Serv 2008 Nov;59(11):1264-1272 [FREE Full text] [doi: 10.1176/appi.ps.59.11.1264] [Medline: 18971402]
3. Chow EA, Foster H, Gonzalez V, McIver L. The disparate impact of diabetes on racial/ethnic minority populations. Clin Diabetes 2012 Jul 16;30(3):130-133. [doi: 10.2337/diaclin.30.3.130]
4. Katon WJ, Rutter C, Simon G, Lin EH, Ludman E, Ciechanowski P, et al. The association of comorbid depression with mortality in patients with type 2 diabetes. Diabetes Care 2005 Nov;28(11):2668-2672. [Medline: 16249537]
5. Colon E, Giachello A, McIver L, Pacheco G, Vela L. Diabetes and depression in the Hispanic/Latino community. Clin Diabetes 2013 Jan 17;31(1):43-45. [doi: 10.2337/diaclin.31.1.43]
6. Aguilera A, Lyles CR. The case for jointly targeting diabetes and depression among vulnerable patients using digital technology. JMIR Diabetes 2017 Jan 17;2(1):e1 [FREE Full text] [doi: 10.2196/diabetes.6916] [Medline: 30291080]
7. Craft LL, Perna FM. The benefits of exercise for the clinically depressed. Prim Care Companion J Clin Psychiatry 2004;6(3):104-111 [FREE Full text] [Medline: 15361924]
8. Silveira H, Moraes H, Oliveira N, Coutinho ES, Laks J, Deslandes A. Physical exercise and clinically depressed patients: a systematic review and meta-analysis. Neuropsychobiology 2013;67(2):61-68 [FREE Full text] [doi: 10.1159/000345160] [Medline: 23295766]
9. Hamasaki H. Daily physical activity and type 2 diabetes: a review. World J Diabetes 2016 Jun 25;7(12):243-251 [FREE Full text] [doi: 10.4239/wjd.v7.i12.243] [Medline: 27350847]
10. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
11. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 07;18(1):e5 [FREE Full text] [doi: 10.2196/jmir.5235] [Medline: 26742999]
12. Ferrer-Roca O, Cárdenas A, Diaz-Cardama A, Pulido P. Mobile phone text messaging in the management of diabetes. J Telemed Telecare 2004;10(5):282-285. [doi: 10.1258/1357633042026341] [Medline: 15494086]
13. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
14. Fortmann AL, Gallo LC, Garcia MI, Taleb M, Euyoque JA, Clark T, et al. Dulce Digital: an mHealth SMS-based intervention improves glycemic control in Hispanics with type 2 diabetes. Diabetes Care 2017 Oct;40(10):1349-1355. [doi: 10.2337/dc17-0230] [Medline: 28600309]
15. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-01-05]
16. 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]
17. Fitts WJ, Furberg R. Underdeveloped or underreported? Coverage of pretesting practices and recommendations for design of text message-based health behavior change interventions. J Health Commun 2015 Apr;20(4):472-478. [doi: 10.1080/10810730.2014.977468] [Medline: 25749250]
18. Parra S, Ortega A, Kanter R, Kain J. Process of developing text messages on healthy eating and physical activity for Chilean mothers with overweight or obese preschool children to be delivered via WhatsApp. Cogent Social Sciences 2018 Oct 08;4(1):1-11. [doi: 10.1080/23311886.2018.1521057]
19. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
20. Shaw RJ, Bosworth HB, Hess JC, Silva SG, Lipkus IM, Davis LL, et al. Development of a theoretically driven mHealth text messaging application for sustaining recent weight loss. JMIR Mhealth Uhealth 2013;1(1):e5 [FREE Full text] [doi: 10.2196/mhealth.2343] [Medline: 25100678]
21. Ybarra ML, Holtrop JS, Bağci BA, Emri S. Design considerations in developing a text messaging program aimed at smoking cessation. J Med Internet Res 2012;14(4):e103 [FREE Full text] [doi: 10.2196/jmir.2061] [Medline: 22832182]
22. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-754. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
23. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]
24. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
25. Nundy S, Dick JJ, Goddu AP, Hogan P, Lu CE, Solomon MC, et al. Using mobile health to support the chronic care model: developing an institutional initiative. Int J Telemed Appl 2012;2012:871925 [FREE Full text] [doi: 10.1155/2012/871925] [Medline: 23304135]
26. 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]
27. Owens C, Farrand P, Darvill R, Emmens T, Hewis E, Aitken P. Involving service users in intervention design: a participatory approach to developing a text-messaging intervention to reduce repetition of self-harm. Health Expect 2011 Sep;14(3):285-295. [doi: 10.1111/j.1369-7625.2010.00623.x] [Medline: 20860777]
28. Mao J, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar 01;48(3):105-109. [doi: 10.1145/1047671.1047677]
29. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
30. Avila-Garcia P, Hernandez-Ramos R, Nouri SS, Cemballi A, Sarkar U, Lyles CR, et al. Engaging users in the design of an mHealth, text message-based intervention to increase physical activity at a safety-net health care system. JAMIA Open 2019 Dec;2(4):489-497 [FREE Full text] [doi: 10.1093/jamiaopen/ooz049] [Medline: 32025646]
31. Diabetes and Depression Text Messaging Intervention (DIAMANTE). ClinicalTrials.gov. 2018. URL: https://clinicaltrials.gov/ct2/show/record/NCT03490253?view=record [accessed 2020-01-20]
32. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
33. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
34. Evans MB, Cooke LM, Murray RA, Wilson AE. The sooner, the better: exercise outcome proximity and intrinsic motivation. Appl Psychol Health Well Being 2014 Nov;6(3):347-361. [doi: 10.1111/aphw.12032] [Medline: 25209956]
35. White KM, Terry DJ, Troup C, Rempel LA. Behavioral, normative and control beliefs underlying low-fat dietary and regular physical activity behaviors for adults diagnosed with type 2 diabetes and/or cardiovascular disease. Psychol Health Med 2007 Aug;12(4):485-494. [doi: 10.1080/13548500601089932] [Medline: 17620212]
36. Rothman A, Bartels R, Wlaschin J, Salovey P. The strategic use of gain- and loss-framed messages to promote healthy behavior: how theory can inform practice. J Communication 2006;56(s1):202-220. [doi: 10.1111/j.1460-2466.2006.00290.x]
37. Mays D, Turner MM, Zhao X, Evans WD, Luta G, Tercyak KP. Framing pictorial cigarette warning labels to motivate young smokers to quit. Nicotine Tob Res 2015 Jul;17(7):769-775 [FREE Full text] [doi: 10.1093/ntr/ntu164] [Medline: 25143295]

38. Larsson LS, Champine D, Hoyt D, Lin L, Salois E, Silvas S, et al. Social marketing risk-framing approaches for dental sealants in rural American Indian children. Public Health Nurs 2015;32(6):662-670 [FREE Full text] [doi: 10.1111/phn.12203] [Medline: 26032902]
39. Farris KB, Salgado TM, Batra P, Piette JD, Singh S, Guhad A, et al. Confirming the theoretical structure of expert-developed text messages to improve adherence to anti-hypertensive medications. Res Social Adm Pharm 2016 Jul;12(4):578-591 [FREE Full text] [doi: 10.1016/j.sapharm.2015.09.009] [Medline: 26525857]
40. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: mixed methods study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
41. DiMatteo MR, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: a three-factor model to guide practice. Health Psychol Rev 2012 Mar;6(1):74-91. [doi: 10.1080/17437199.2010.537592]
42. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
43. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006 Jan;60(1):7-12 [FREE Full text] [doi: 10.1136/jech.2004.023531] [Medline: 16361448]
44. Korkiakangas EE, Alahuhta MA, Laitinen JH. Barriers to regular exercise among adults at high risk or diagnosed with type 2 diabetes: a systematic review. Health Promot Int 2009 Dec;24(4):416-427. [doi: 10.1093/heapro/dap031] [Medline: 19793763]
45. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
46. Readable. 2021. URL: https://readable.com/ [accessed 2019-04-09]
47. Buchholz SW, Sandi G, Ingram D, Welch MJ, Ocampo EV. Bilingual text messaging translation: translating text messages from English into Spanish for the Text4Walking program. JMIR Res Protoc 2015 May 06;4(2):e51 [FREE Full text] [doi: 10.2196/resprot.3984] [Medline: 25947953]
48. González AM. Español neutro — español latino: hacia una norma hispanoamericana en los medios de comunicación. Roczniki Humanistyczne 2019 Jul 24;67(5):7. [doi: 10.18290/rh.2019.67.5-1]
49. Zickar MJ. Measurement development and evaluation. Annu Rev Organ Psychol Organ Behav 2020 Jan 21;7(1):213-232. [doi: 10.1146/annurev-orgpsych-012119-044957]

Abbreviations

COM-B: capability, opportunity, motivation, and behavior
DIAMANTE: Diabetes and Mental Health Adaptive Notification Tracking and Evaluation
mHealth: mobile health
MTurk: Amazon Mechanical Turk
UCD: user-centered design

Edited by L Buis; submitted 09.06.20; peer-reviewed by C Stiles-Shields, S Guo; comments to author 27.07.20; revised version received 21.09.20; accepted 13.04.21; published 19.05.21.

Please cite as:
Pathak LE, Aguilera A, Williams JJ, Lyles CR, Hernandez-Ramos R, Miramontes J, Cemballi AG, Figueroa CA
Developing Messaging Content for a Physical Activity Smartphone App Tailored to Low-Income Patients: User-Centered Design and Crowdsourcing Approach
JMIR Mhealth Uhealth 2021;9(5):e21177
URL: https://mhealth.jmir.org/2021/5/e21177
doi:10.2196/21177
PMID:34009130

Original Paper

Comparing the Efficacies of Telemedicine and Standard Prenatal Care on Blood Glucose Control in Women With Gestational Diabetes Mellitus: Randomized Controlled Trial

Ying Tian[1], BA; Suhan Zhang[1], BA; Feiling Huang[1], BA; Liangkun Ma[1], MD, PhD

Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Corresponding Author:
Liangkun Ma, MD, PhD
Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric & Gynecologic Diseases
Peking Union Medical College Hospital
Chinese Academy of Medical Sciences & Peking Union Medical College
Peking Union Medical College Hospital (Dongdan campus)
No 1 Shuaifuyuan, Wangfujing Dongcheng District
Beijing, 100730
China
Phone: 86 13021961166
Email: MaLiangKun@pumch.cn

Abstract

Background: Gestational diabetes mellitus (GDM) can usually be well controlled by health education and lifestyle management, resulting in better pregnancy outcomes. However, standard clinical prenatal care, which consists of clinic visits every 2 weeks, may not provide sufficient management for women with GDM. Telemedicine demonstrates a potential to fill this gap.

Objective: The objective of this study was to investigate whether health education and lifestyle management delivered through a WeChat group chat was more effective in controlling blood glucose (BG) than standard clinical prenatal care among women with GDM.

Methods: In this multicenter randomized controlled trial, women with GDM diagnosed by an oral glucose tolerance test between 23 and 30 (+6) gestational weeks were randomized to a WeChat group chat–based BG management group or a routine clinical prenatal care group. The primary outcome was the change in the glycemic qualification rate during the follow-up period in both groups. The secondary outcomes were pregnancy outcomes.

Results: A total of 309 women with GDM participated in the trial, with 162 women randomized to the control group and 147 to the intervention group. No significant differences in baseline characteristics were found between the control and intervention groups. Participants were further divided into 4 groups according to gestational weeks at enrollment for further analysis. The glycemic qualification rate of the intervention group was higher than that of the control group at nearly all time points in Groups 1 to 3, among which 3 time points reached statistical significance: Group 1 at T3 (54.8% vs 83.3%) and Group 2 at T3 (62.5% vs 80.0%) and T7 (75.0% vs 100%). The glycemic qualification rate gradually increased as gestational weeks progressed in both groups, regardless of the intervention method. None of the pregnancy outcomes measured, including delivery mode, premature rupture of the membranes, preterm birth, infant's birth weight, and postpartum hemorrhage, were significantly different between the control and intervention groups.

Conclusions: This multicenter randomized controlled trial that assessed women with noninsulin-dependent GDM demonstrated that additional instant messaging platforms, such as WeChat, used for health education and lifestyle intervention in China tend to be more effective for BG control than standard clinical prenatal care alone.

Trial Registration: ClinicalTrials.gov NCT03748576; https://clinicaltrials.gov/ct2/show/NCT03748576

(JMIR Mhealth Uhealth 2021;9(5):e22881) doi:10.2196/22881

KEYWORDS

gestational diabetes mellitus; blood glucose; mhealth; health education; lifestyle management

Introduction

Gestational diabetes mellitus (GDM) has been defined as carbohydrate intolerance resulting in hyperglycemia of variable severity with an onset or first recognition during pregnancy that resolves after delivery [1]. Although this condition of carbohydrate intolerance is maintained only during the peripartum period, the potential risks are not limited to pregnancy outcomes. Short-term impacts include a higher risk of developing gestational hypertension and preeclampsia; an increased risk of preterm birth, cesarean section, perineal trauma, and postpartum hemorrhage for mothers [2,3]; and an increased risk of macrosomia, large for gestational age status, birth trauma, and neonatal hypoglycemia for infants [4]. Potential long-term effects include but are not limited to a risk of developing type 2 diabetes within 5 to 10 years, higher risks of developing cardiovascular disease for mothers [5], and higher rates of obesity, late-onset diabetes, and cardiovascular disease in adulthood for infants [6-8]. Certain ethnic groups, such as Asians, have a higher risk of GDM [9]. The aim of GDM treatment and patient education is to keep glucose levels within the recommended glycemic reference range to prevent maternal hyper- or hypoglycemia [10,11]. Lifestyle modifications, including diet consultation and physical exercise, are considered the first-line intervention and may suffice for most patients [10]. Women with GDM who are on insulin therapy receive more attention from doctors and researchers, while women with GDM with mild dysglycemia who control their condition with diet alone receive much less attention, although this group may account for the majority of patients with GDM [12].

Patient education—including improving GDM-related knowledge and clarifying the importance of self-management, and multidisciplinary consultation—including blood glucose (BG) monitoring, lifestyle modifications, and drug therapy—are necessary for women with GDM after receiving a GDM diagnosis. Additionally, there are some cell phone applications and WeChat Official Accounts that are specifically for women to use during pregnancy to improve disease knowledge and monitor BG and body weight. Regardless of access to medical advice and self-management, managing and educating women with GDM is not consistent, and the adherence of patients is unsupervised in most cases. These factors may lead to consequences, such as reducing GDM management efficiency and causing poor pregnancy outcomes [13]. It was reported that good adherence of patients with GDM to clinical recommendations was observed, especially in relation to BG monitoring, but nutritional guidelines were less likely to be followed [14]. Given the limited intervention time window, the importance of continuous patient education and self-management supervision during the periods between antenatal care visits should be strongly emphasized.

The rapid development of the internet and smartphones has caused telemedicine to be more convenient. For women with GDM, online management can potentially improve the likelihood of following medical advice and keep patients with GDM alert to the harm of poor BG control, further achieving continuous disease management. Research has demonstrated the acceptability and feasibility of mobile medical management of women with GDM through social software platforms, instant messaging applications, web pages, and regular email alerts or telephone visits. Some of the related studies did show a slight correlation between mobile health applications and BG levels, as well as better efficacy and greater demand for more attention, communication, and management from a majority of women with GDM [15-17]. However, to date, telemedicine approaches have failed to show significant advantages over standard prenatal care, especially in terms of clinical outcomes, when used for GDM. Possibly because of the limited sample sizes and immature intervention paths, it is difficult to identify whether a relationship between technology use and clinical outcomes exists [15-23]. A large-sample randomized controlled study is needed to further explore the application value of mobile medicine regarding BG control, health management, and clinical outcomes.

WeChat, an instant messaging platform with more than 1.13 billion monthly global active users, is widely used not only for message delivery but also gradually in the medical field, including but not limited to online consultation, patient group management, and live patient education. The potential of WeChat for telemedicine applications cannot be ignored. We designed a WeChat group chat–based GDM management program to conduct health education and lifestyle management online to facilitate continuous patient education and self-management during the periods between prenatal care visits. The patient management team included obstetricians, nutritionists, nurses, and health managers at a minimum, and psychologists and sports medicine doctors were included if needed. Based on our clinical experience and patients' responses, we hypothesized that health education and lifestyle management offered by medical experts on WeChat would result in better control of BG and better pregnancy outcomes. This randomized controlled trial was conducted to verify this hypothesis by comparing the efficacy of health education and lifestyle management conducted through the WeChat group chat on BG control with that of standard clinical prenatal care in women with GDM.

Methods

Study Design

This was a multicenter randomized controlled trial initiated by the Chinese Academy of Medical Sciences and Peking Union Medical College Hospital and conducted in 8 prenatal care institutions between June 2019 and December 2019. The trial received ethical approval from the Peking Union Medical College Hospital's Medical Ethics Review Committee (ID: JS-1012). Written informed consent was obtained from each participant. All researchers involved had been trained uniformly before the trial started, and all centers received site instruction when they enrolled the first several participants.

Participants

The participant inclusion criteria were as follows: (1) aged between 18 and 45 years; (2) singleton pregnancy of fewer than 31 gestational weeks; (3) GDM diagnosis according to the 75 g oral glucose tolerance test (OGTT) [24] and no requirement for insulin treatment according to multidisciplinary consultation;

(4) ability to use a smartphone for chatting, reading, and writing basic Chinese; and (5) voluntary participation in research. Pregnant women with a diagnosed chronic disease and other pregnancy complications except GDM, as well as those who had recent trauma or treatment with glucocorticoids, were excluded.

Eligible patients according to assessment by clinic doctors based on the abovementioned criteria received detailed information about the project from the research team. Interested patients completed questionnaires after providing signed informed consent and were randomized to a study group. During the study, all subjects could withdraw at any time without providing a reason. Participants were informed that they would receive postpartum counseling for free as a reward for participating in the study.

Sample Size

The sample size for this superiority trial was calculated as Inline graphic 1, where P_C was the BG control rate in the control group and P_T was the expected BG control rate in the test group (referring to clinical data). The absence of published estimates of glycemic control rates made sample size calculations more challenging. We decided to assume that the control rate in the control group was approximately 75%, based on our clinical experience, and that the control rate in the experimental group would be 15% higher than that in the control group. Based on these assumptions, we calculated the sample size to be 104 cases in each group at the significance level of .05 (bilateral) and 80% assurance. Considering a dropout rate of up to 20%, the final sample size was determined to be 125 women in each group.

Allocation

Participants were randomly allocated to two groups: intervention (WeChat group management) or control (standard clinical prenatal care). A random number table was used to generate the grouping envelope. Participants were grouped according to the random number in the grouping envelope, ensuring randomized allocation. After the results of the grouping were announced, the subjects were not permitted to switch groups.

Intervention and Control

Standard Clinical Prenatal Care (Control Group)

All women diagnosed with GDM were asked to attend one GDM management lesson in a maternity school organized by prenatal care institutions, which were established according to the standards of the Beijing Municipal Health Commission. Participants were taught basic information about GDM and how to do self-management, including how to conduct BG monitoring, what the target BG values are, and how to keep a lifestyle diary. The routine prenatal care appointments were changed to once every 2 weeks when GDM was diagnosed. Doctors generally asked women with GDM to record 5 BG values daily—fasting and pre-sleep BG, and 2-hour postprandial BG (postbreakfast, postlunch, and postdinner BG)—and schedule clinic visits with at least 3 days between visits. At clinic visits, doctors checked the details of patients' lifestyle diaries, including daily diet, exercise, weight, BG, and blood pressure, and lifestyle guidance was provided based on their records. The guidance involved simple principles of a healthy lifestyle such as increasing food diversity, consuming high-fiber cereals, avoiding eating outside, and taking more exercise for weight control. If patients failed to present their diaries, doctors asked them to return with the records on the following week. If the BG could not be controlled after general lifestyle guidance, drug-based interventions were considered after multidisciplinary consultation.

WeChat Group Management (Intervention Group)

Patients in the intervention group received WeChat group management in addition to maternity school and standard clinical prenatal care. In the WeChat group chat, participants received management on a weekly basis. In particular, every Monday, researchers would issue a briefing to encourage patients to take an active part in the control of their GDM and a task card to pinpoint the basic requirements, including diet advice, examples of meals from other group members, and exercise rules. Patients performed self-management according to the basic criteria provided for their actual situation and shared photos of their meals and snacks, daily exercise, and experience regarding BG control. Researchers would give individualized guidance for self-management or use a group member's situation as an example for others. In this way, participants could learn not only from their personalized guidance but also from the experiences of other group members. On weekends, the researchers prepared lessons and articles for group members to learn different aspects of pregnancy and GDM, including rudimentary knowledge, disease management, psychology, and past cases. We encouraged the sharing of learning experiences and notes in the form of peer interactions and support groups. If there were any questions regarding the project, pregnancy, or GDM, patients could seek answers from the group chat. This weekly management continued until delivery.

Follow-Up Plan

Participants visited the obstetric clinic for prenatal care and follow-up every 2 weeks beginning at the time of enrollment, and each 2-week period was referred to as a "T." We asked the participants in both groups to record their fasting, postbreakfast, postlunch, postdinner, and pre-sleep BG values for 6 days in 1 T. In other words, within a T, we obtained 30 monitored values. For example, if someone enrolled at 24 gestational weeks and delivered at 40 gestational weeks, she would provide 8 follow-up records; 24 gestational weeks would be recorded as T0, and the follow-up records would be recorded from T1 to T8. Conversely, if a participant enrolled at 30 gestational weeks and delivered at 38 gestational weeks, she would have only 4 follow-up records, with T0 as the enrollment visit and T1 to T4 as the follow-up period.

Statistical Analysis

The primary outcome in the study was the glycemic qualification rate. The glycemic qualification rate was calculated as the number of BG levels within the control range/30 × 100%. The BG control ranges were fasting BG (fasting and pre-sleep BG) <95 mg/dL (5.3 mmol/L) and 2-hour postprandial BG (postbreakfast, postlunch, and postdinner BG) <120 mg/dL (6.7 mmol/L) [25]. Secondary outcomes including delivery mode,

premature rupture of the membranes, preterm birth, infant's birth weight, and postpartum hemorrhage were compared between the study groups.

Data analysis was carried out using SAS software, version 9.4 (SAS Institute Inc). The descriptive statistics of continuous data are expressed as mean (SD). For comparisons between the groups, the *t* test was used if both groups fulfilled the criteria of the test of normality and homogeneity of variance; otherwise, the nonparameter Wilcoxon rank sum test was used. Descriptive statistics are expressed as n (%), and differences in categorical variables were assessed using the chi-square test. Boxplots were used to illustrate changes in BG during pregnancy. Differences in outcomes between the 2 groups were compared using chi-square analysis. $P<0.05$ was considered statistically significant.

Results

Participant Characteristics

As shown in Figure 1, 365 women were screened when they came to the prenatal clinic to check OGTT results; 56 of them were excluded because they did not meet the inclusion criteria (n=22) or they declined to participate (n=34). Of 309 women with GDM who met the recruitment criteria and signed informed consent forms, 162 were randomized to the control group and 147 were randomized to the intervention group. Eleven women in the control group and 6 women in the intervention group did not record any BG values. Fifteen women in the control group and 8 women in the intervention group changed their delivery hospital; thus, delivery records were not available. None of the participants switched to insulin therapy during the program because of persistently severe glucose abnormality or poor pregnancy outcome. Data from 136 women in the control group and 133 in the intervention group were included in the analysis.

No significant differences were found between the control and intervention groups at baseline (Table 1; and Multimedia Appendix 1, Tables 1-4). Of the 269 participants in the analysis, 56 (20.8%) women were over age 35 years, 71 (26.4%) were overweight, and 37 (13.8%) were obese. The OGTT results at enrollment were as follows: fasting BG 4.98 (SD 0.78) mmol/L, 1-hour BG 10.20 (SD 1.84) mmol/L, and 2-hour BG 8.89 (SD 1.69) mmol/L. Participants could be further divided into 4 groups according to gestational weeks at enrollment: Group 1, 66 women who enrolled between 23 and 24 (+6) gestational weeks; Group 2, 113 women who enrolled between 25 and 26 (+6) gestational weeks; Group 3, 66 women who enrolled between 27 and 28 (+6) gestational weeks; and Group 4, 24 women enrolled between 29 and 30 (+6) gestational weeks (n=24).

Figure 1. Study flowchart.

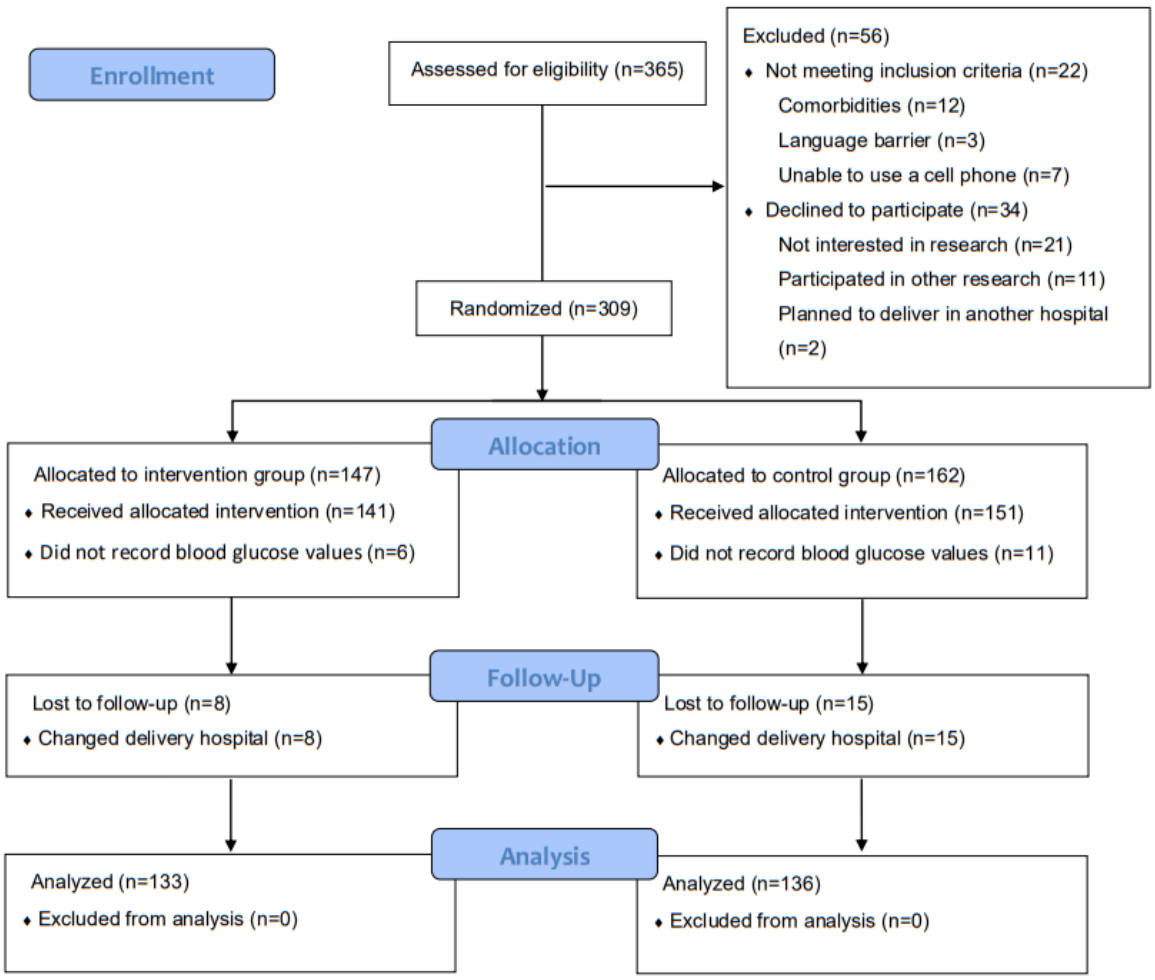


Table 1. Baseline characteristics of the intervention and control groups.

Variable	Total (N=269)	Control group (n=136)	Intervention group (n=133)	Wilcoxon rank sum or χ^2 test	*P* value
Age (years), mean (SD)	31.07 (4.34)	30.93 (4.48)	31.23 (4.21)	0.95[a]	.34
Age (years), n (%)				0.04[b]	.84
≤35	213 (79.1)	107 (78.7)	106 (79.7)		
>35	56 (20.8)	29 (21.3)	27 (20.3)		
BMI, mean (SD)	24.04 (11.43)	25.07 (15.62)	22.99 (3.64)	–0.95[a]	.34
BMI (kg/m^2) , n (%)				2.85[b]	.41
10-18.5	24 (8.9)	12 (8.8)	12 (9.0)		
18.5-24	137 (50.9)	69 (50.7)	68 (51.1)		
24-28	71 (26.4)	32 (23.5)	39 (29.3)		
≥28	37 (13.8)	23 (16.9)	14 (10.5)		
Gravidity, n (%)				0.04[b]	.84
Primigravida	123 (45.7)	63 (46.3)	60 (45.1)		
Multigravida	146 (54.3)	73 (53.7)	73 (54.9)		
Parity, n (%)				0.42[b]	.52
Primipara	161 (59.9)	84 (61.8)	77 (57.9)		
Multipara	108 (40.1)	52 (38.2)	56 (42.1)		
Newborn sex, n (%)				0.45[b]	.50
Boy	134 (49.8)	65 (47.8)	69 (51.9)		
Girl	135 (50.2)	71 (52.2)	64 (48.1)		
Nationality, n (%)				1.76[b]	.19
Ethnic Han	259 (96.3)	133 (97.8)	126 (94.7)		
Other	10 (3.7)	3 (2.2)	7 (5.3)		
Family history of diabetes mellitus, n (%)				1.30[b]	.25
Yes	12 (4.5)	8 (5.9)	4 (3.0)		
No	257 (95.5)	128 (94.1)	129 (97.0)		
OGTT[c]values (mmol/L), mean (SD)					
Fasting	4.98 (0.78)	5.04 (0.75)	4.93 (0.81)	–1.56[a]	.12
1 hour	10.20 (1.84)	10.03 (1.88)	10.37 (1.79)	0.87[a]	.39
2 hour	8.89 (1.69)	8.83 (1.63)	8.96 (1.75)	–0.03[a]	.97
Gestational weeks at enrollment, mean (SD)	26.37 (1.74)	26.27 (1.72)	26.48 (1.77)	1.14[a]	.26
Patient groups according to gestational weeks at enrollment, n (%)				1.71[b]	.63
Group 1	66 (24.5)	34 (25.0)	32 (24.1)		
Group 2	113 (42.0)	61 (44.9)	52 (39.1)		
Group 3	66 (24.5)	29 (21.3)	37 (27.8)		
Group 4	24 (8.9)	12 (8.8)	12 (9.0)		

[a]Wilcoxon rank sum test.

[b]Chi-square test.

[c]OGTT: oral glucose tolerance test.

Primary Outcome

The glycemic qualification rate of the intervention group was higher than that of the control group at nearly all time points in Groups 1 to 3, with a statistically significant difference at 3 time points: Group 1 at T3 (54.8% vs 83.3%) and Group 2 at T3 (62.5% vs 80.0%) and T7 (75.0% vs 100%) (Table 2).

Table 2. Glycemic qualification rate of the control group (CG) and intervention group (IG) according to gestational weeks at enrollment.

Time point	Glycemic qualification rate (%), median (Q1, Q3)							
	Group 1		Group 2		Group 3		Group 4	
	CG	IG	CG	IG	CG	IG	CG	IG
T1	59.17 (40.00, 80.00)	60.77 (40.00, 80.00)	60.00 (40.00, 80.00)	66.67 (41.67, 81.82)	50.00 (25.00, 61.90)	50.00 (50.00, 83.33)	48.33 (37.50, 73.33)	50.00 (33.33, 62.50)
T2	53.33 (40.00, 88.89)	75.00 (50.00, 100)	62.50 (37.50, 87.50)	70.00 (52.94, 90.00)	62.50 (40.00, 100)	61.25 (50.00, 100)	45.00 (30.00, 74.17)	62.50 (40.00, 100)
T3	54.79 (46.67, 80.00)	83.33 (62.50, 100)*	62.50 (50.00, 75.00)	80.00 (62.50, 87.50)*	62.50 (50.00, 90.00)	80.00 (50.00, 100)	52.78 (40.00, 87.50)	50.00 (26.67, 75.00)
T4	80.00 (50.00, 100)	100 (73.33, 100)	70.59 (60.00, 100)	80.00 (62.50, 100)	75.00 (53.33, 100)	75.00 (53.33, 100)	70.00 (T50.00, 100)	63.33 (57.14, 100)
T5	70.00 (50.00, 80.00)	87.50 (66.67, 100)	87.50 (62.50, 100)	80.00 (75.00, 100)	87.50 (62.50, 100)	87.50 (60.00, 100)	88.75 (73.75, 95.00)	62.50 (45.00, 82.50)
T6	87.50 (46.67, 100)	83.33 (57.35, 93.75)	75.00 (50.00, 100)	81.67 (75.00, 100)	75.00 (55.00, 100)	83.89 (50.00, 100)	75.00 (50.00, 80.00)	60.00 (60.00, 60.00)
T7	80.00 (70.00, 87.50)	88.89 (75.00, 100)	75.00 (60.00, 87.50)	100 (87.50, 100)*	75.00 (62.50, 81.25)	75.00 (60.00, 87.50)	–	–
T8	87.50 (40.00, 100)	100 (60.00, 100)	87.50 (60.00, 100)	90.97 (80.00, 100)	62.50 (62.50, 62.50)	100 (100, 100)	–	–
T9	50.00 (50.00, 50.00)	80.00 (80.00, 80.00)	75.00 (50.00, 100)	100 (100, 100)	–	–	–	–

*Statistically significant at $P<.05$.

With respect to intervention period, at T1, when participants began to receive health management, the BG qualification rates of the control and intervention groups were similar. At T2, the BG qualification rates in the control group and the intervention group both began to rise (53.3% vs 75.0%, respectively, in Group 1 and 62.5% vs 70.0%, respectively, in Group 2). At T3, when the intervention group had been managed for approximately 1 month, the BG qualification rates in the control and intervention groups were 54.8% vs 83.3%, respectively, in Group 1, 62.5% vs 80.0%, respectively, in Group 2, and 62.5% vs 80.0%, respectively, in Group 3.

With regard to gestational week of enrollment, the earlier the intervention started, the better BG was controlled. That is, in Group 1, who enrolled between 23 and 24 (+6) gestational weeks, the control rate was not different between the two group at T1, but once the intervention started, the BG control rate of the intervention group became much higher than that of the control group (75.0% vs 53.3%) at T2. The difference was seen until T5, and then decreased to less than 10% after that. In Group 2, who enrolled between 25 and 26 (+6) gestational weeks, the trend was similar, although the difference was only seen until T4 and then decreased. However, in Group 4, a difference in the BG control rate between the intervention and control groups was not observed.

Furthermore, the glycemic qualification rate gradually increased as gestational weeks progressed in both groups, regardless of the intervention or control group assignment (Table 2 and Figures 2-5).

Figure 2. Glycemic qualification rate of the intervention and control groups in Group 1 (enrolled between 23 and 24 [+6] gestational weeks) at different time points. T: 2-week period.

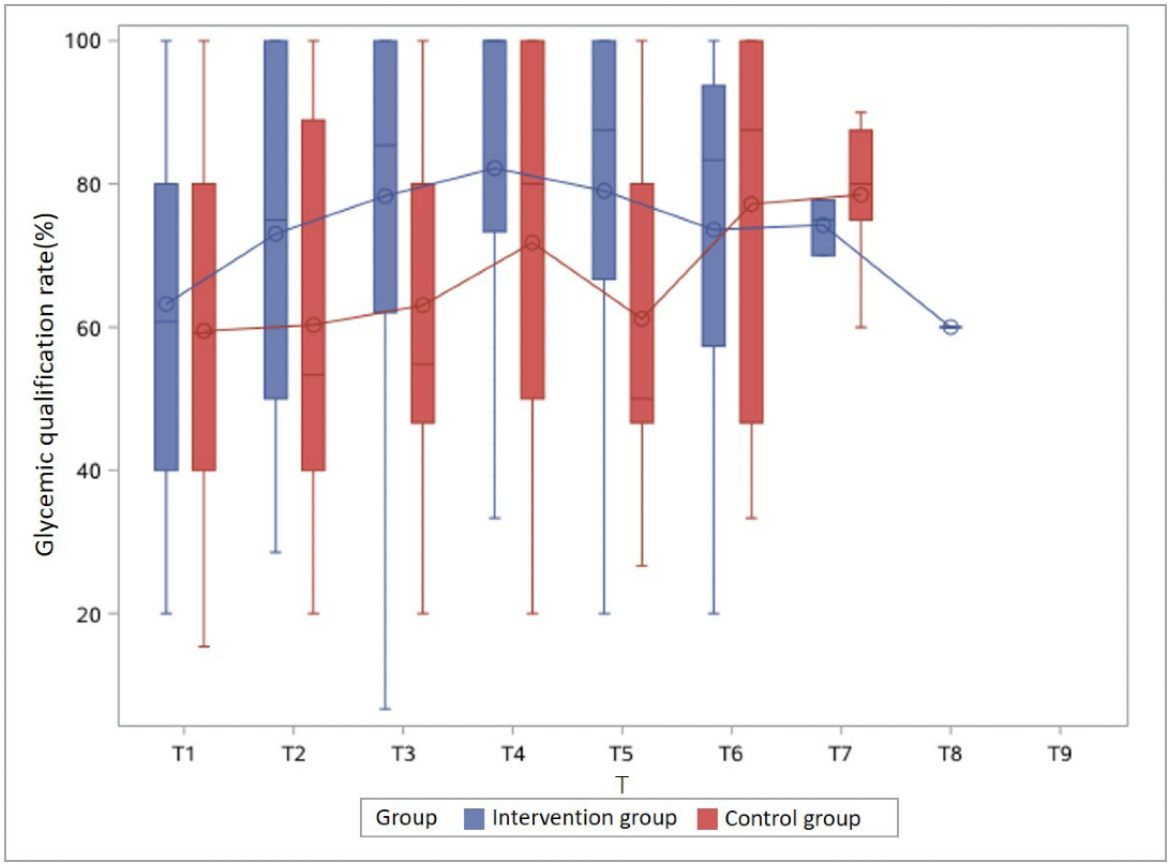


Figure 3. Glycemic qualification rate of the intervention and control groups in Group 2 (enrolled between 25 and 26 [+6] gestational weeks) at different time points. T: 2-week period.

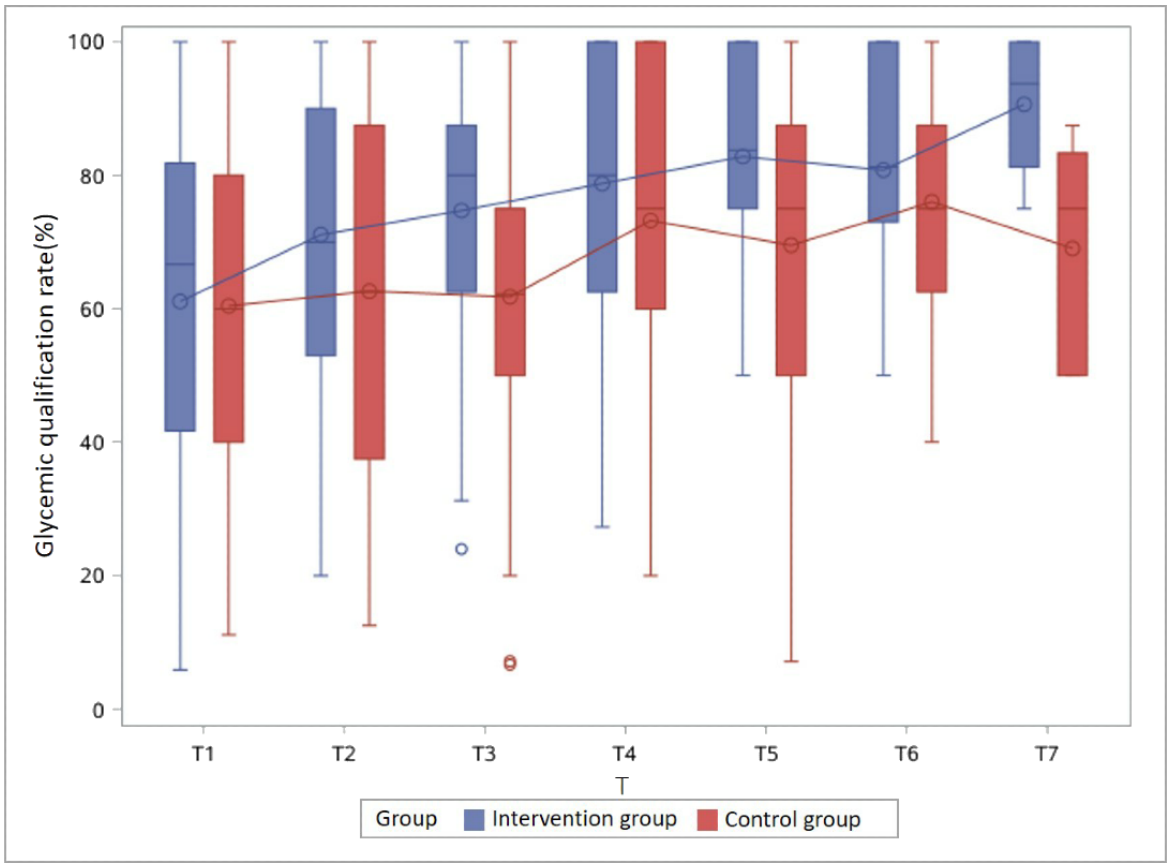


Figure 4. Glycemic qualification rate of the intervention and control groups in Group 3 (enrolled between 27 and 28 [+6] gestational weeks) at different time points. T: 2-week period.

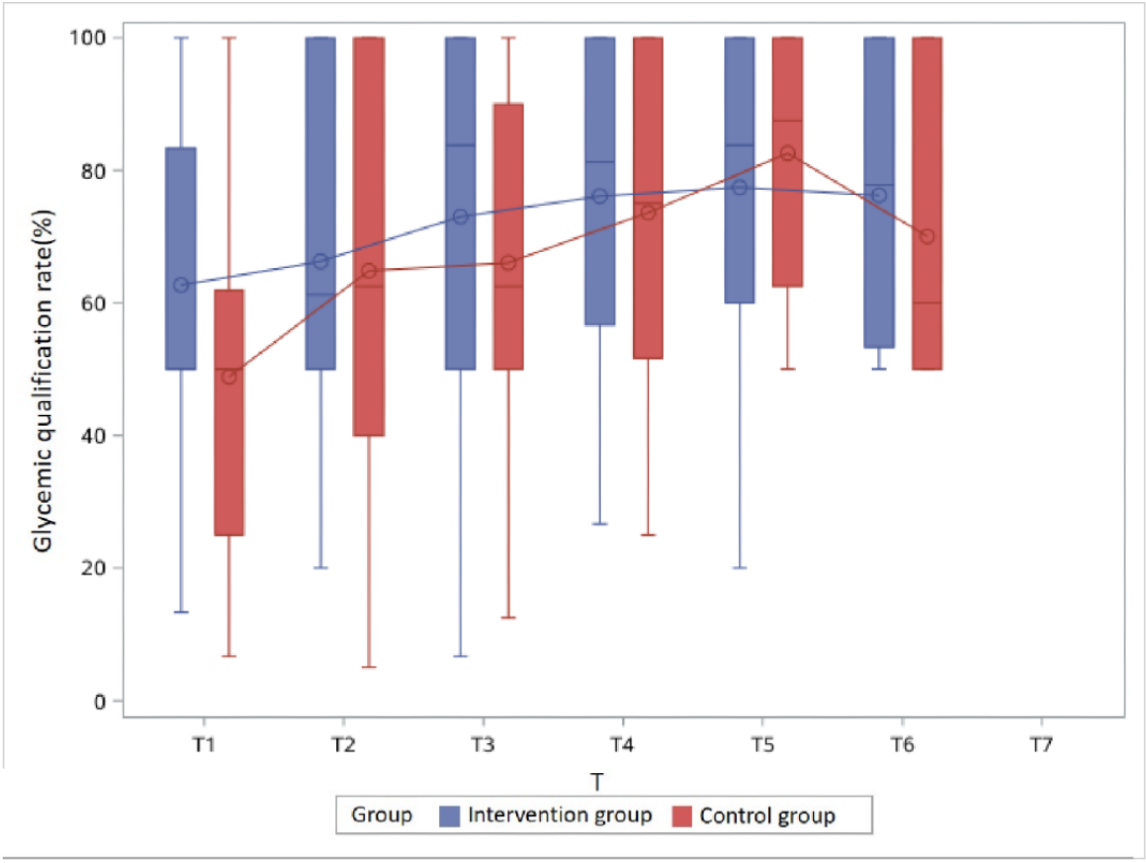
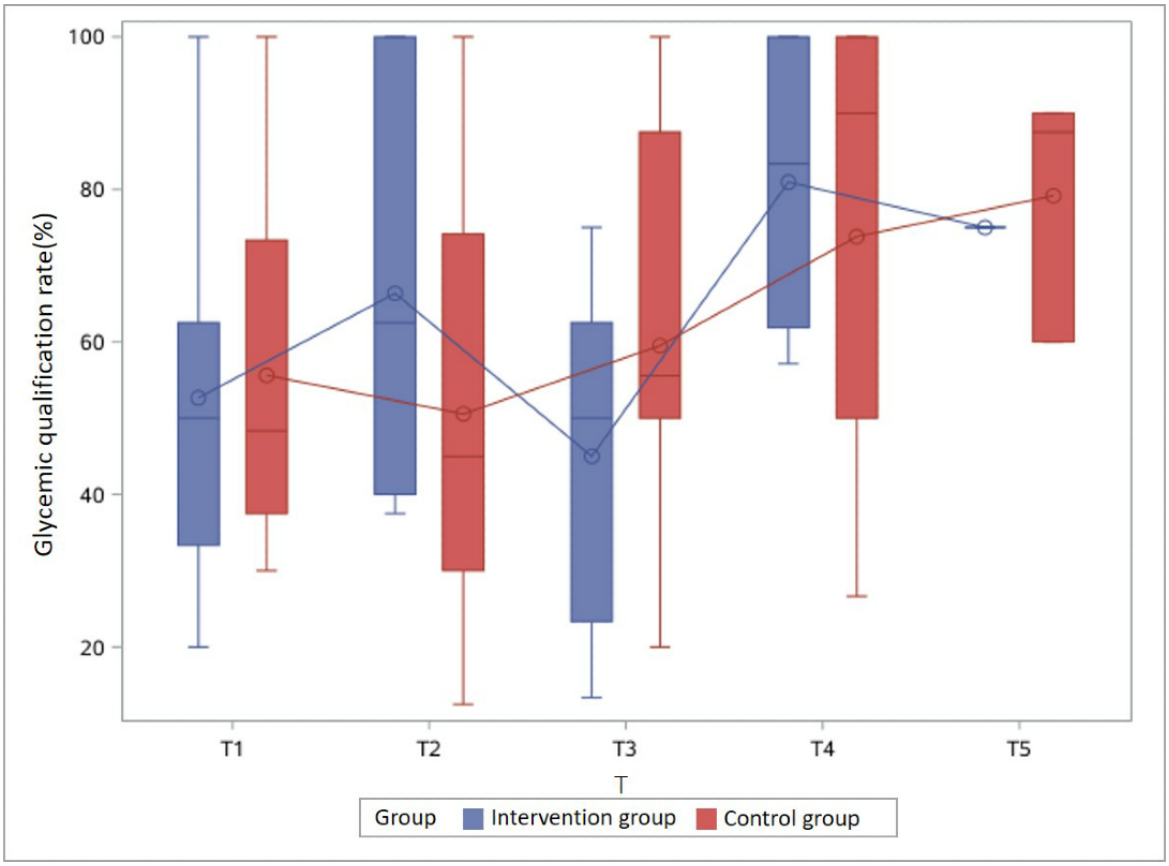


Figure 5. Glycemic qualification rate of the intervention and control groups in Group 4 (enrolled between 29 and 30 [+6] gestational weeks) at different time points. T: 2-week period.



Secondary Outcomes

Secondary outcomes including delivery mode, premature rupture of the membranes, preterm birth, infant's birth weight, and postpartum hemorrhage were not significantly different between the control and intervention groups (see Multimedia Appendix 1, Tables 5-8).

Discussion

Principal Findings

This randomized controlled trial indicated that the use of instant messaging platforms such as WeChat in China for health education and lifestyle intervention by women with GDM for the entire GDM management period was associated with a higher rate of BG falling within the optimal range than with standard clinical prenatal care, although these results were not statistically significant. Clinical pregnancy outcomes were not significantly different between the intervention and control groups. Additionally, the increasing trend in the glycemic qualification rate was seemingly more stable in the WeChat management group than in the control group, which supports the nonnegligible role of continuous management. As well, GDM management was beneficial for improving the BG qualification rate in both groups regardless of when patients enrolled in the study, which means that when the intervention starts earlier, optimal BG is maintained longer and maternal benefits are more extensive. This highlights the importance of starting health management when GDM is diagnosed.

Comparison with Previous Studies

Telemedicine, concentrating on the intervals between obstetric clinic visits, makes GDM management more comprehensive and complete in a limited time window. This can make pregnant women who are unconcerned or overwhelmed by guilt and anxiety feel more confident to manage their GDM. WeChat group chat provides a platform for effective education to ensure clear understanding of medical principles and to provide active pregnancy consultations instead of women having to wait for clinic visits every few weeks. This continuous BG and lifestyle management provides more individualized and adequate practical support than do doctors' recommendations and medical guidelines alone and makes doctors' orders more feasible in practice, which results in better adherence [18,26]. In other words, greater self-efficacy results in better patient management. In addition, among all kinds of patient education and lifestyle interventions, dietary changes are the most difficult recommendations to follow [14]. Factors that impede the implementation of medical recommendations include social activities such as eating out, cultural factors such as the consumption of specific kinds of food, family impacts such as the rejection of the concept of an individual serving size, obsolescent views on pregnancy, and women's family-centered values and role commitments [27,28]. The WeChat group brings together a group of people with specific but similar needs and provides moral support and opportunities for peer communication and family education, making it easier to follow patient management principles. Furthermore, telemedicine has the potential to save time and money by reducing the number of hospital appointments for women with GDM and increase clinical pressure for doctors. This concept is underexplored.

Many studies have explored the effects of online health education and lifestyle management methods on women with GDM. This article is, to our knowledge, the first to interpret BG data from this perspective. Prior studies have indicated that online education and management can help improve BG control [20,29-31]. Our findings confirmed this, although the results of this trial were not statistically significant. The glycemic qualification rates remained higher at nearly all time points regardless of the gestational weeks when participants started receiving online GDM management. This trend became less obvious and stable as the enrollment time drew closer to delivery, which was especially evident in Group 4 and can be attributed to the decreased sample size of participants who had delivered in this group. Some of the previous studies showed no improvement in BG control but better patient satisfaction [18]. Regardless of the management methods, BG control improved during the entire period of pregnancy, as shown in our results and previous research [18]. However, a more detailed glycemic qualification rate revealed that trends did not always improve. We observed that the glycemic qualification rate in the control group did not always increase as it did in the intervention group in Groups 1 to 4. As studies have reported the harm in BG variability for both mothers and infants [32], this discovery emphasizes that continuous management may help decrease BG variability and result in better outcomes. Regarding pregnancy outcomes, a few studies showed relationships between telemedicine interventions and longer gestational periods and fewer preterm births [18], reduced odds of cesarean section and pregnancy-induced hypertension [33], higher Apgar scores and reduced neonatal hospitalization [29]. We failed to investigate these relationships. A common result was that it was difficult to demonstrate differences between groups in pregnancy outcomes [21,23,31]. The same is true for systematic reviews and meta-analyses [34,35]. Compared with previous work, we included an analysis of intervention duration. The glycemic qualification rate increased gradually after GDM management was initiated in both the intervention and control groups regardless of the enrollment time, although the rate in the WeChat group was always higher than that in the control group. It might be interpreted that the earlier that education and lifestyle interventions start, the less time the pregnant woman and fetus are exposed to hyperglycemia and the greater the likelihood of better outcomes.

Strengths and Limitations

The strengths of this study are its multicenter randomized design with a relatively large sample size that allowed us to investigate the effect of both the use of online GDM management and the prolongation of intervention duration. We enrolled women diagnosed with GDM at different gestational weeks and divided them into 4 groups, allowing a more detailed interpretation of the effects of online management on GDM. Moreover, instant messaging platforms such as WeChat may have better generalization than specifically designed webpages or applications because they have fewer requirements for managed patients and require no additional operations except chatting and following the instructions of the managers of the group. In

addition, the high frequency of communication in the chat group and the positive response to health management information indicated that this online management method was welcomed. We conducted brief interviews after the research and obtained positive responses about the intervention.

Limitations also exist in this study. We did not use a continuous glucose monitoring system (CGMS) to obtain BG data. There is a possibility that some glucose information was missed, but studies and guidelines consider that a CGMS is not necessary in women with noninsulin-dependent GDM [36]. Additionally, we failed to demonstrate differences in pregnancy outcomes, as did most previous research, but we could not conclude that the intervention did not have an impact on clinical outcomes. The number of glucose records decreased in the third trimester. We could not determine whether the subjects had monitored their BG but did not record it or whether they had not tested their BG, but the missing values had no effect on the results after assessment.

General management of women with noninsulin-dependent GDM requires lifestyle interventions to realize glycemic control and weight gain management. If we are to implement a more convenient intervention method, comprehensive assessment needs to be done—for example, by assessing its effects on weight change, BG control, sleep quality, and psychological status. An assessment of indicators of adherence is required to confirm the management effect. A larger sample size may help investigate the impact of telemedicine use on the pregnancy outcomes of women with GDM. This investigation is indispensable, as the aim of GDM management is to reduce negative outcomes of the pregnant woman and fetus through BG and weight control. Therefore, the lack of statistically significant results should be further explored in a larger population. A continuous management system for women with GDM to be implemented between prenatal clinics, allowing professional education to be more accessible, is considered to have the potential to reduce clinic visits and improve medical service quality. In the long run, this management method can be extended to GDM prevention at the beginning of pregnancy and GDM follow-up postpartum, emphasizing the continuity of healthy lifestyle education and intervention [37,38].

Conclusions

This multicenter randomized controlled trial that enrolled noninsulin-dependent women with GDM demonstrated that using an instant messaging platform for health education and lifestyle intervention was more effective regarding BG control than standard clinical prenatal care. Further studies may investigate a more systematic method of GDM management conducted through online group chats. Differences in pregnancy outcomes were not significant. More exploration is needed to clarify this trend. These findings also call for more exploitation of telemedicine use for GDM prevention, patient management, and follow-up.

Acknowledgments

The study was supported by Beijing Municipal Science and Technology Commission (grant reference Z161100000516117). The authors sincerely acknowledge the contribution of the women participating in this trial. The authors also acknowledge all researchers working for this trial in every study center, including Peking Union Medical College Hospital, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Shenyang Maternity and Child Health Hospital, QuanZhou Women's and Children's Hospital, Shenyang Women's and Children's Hospital, Xiamen Maternal and Child Health Hospital, Shunyi Women’s & Children’s Hospital of Beijing Children’s Hospital, and Jinan Maternity and Child Care Hospital Affiliated to Shandong First Medical University, Fuyang People's Hospital, and Beijing Pinggu Hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics and pregnancy outcomes of each group.
[DOCX File , 69 KB - mhealth_v9i5e22881_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1152 KB - mhealth_v9i5e22881_app2.pdf]

References

1. World Health Organization. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: a World Health Organization Guideline. Diabetes Res Clin Pract 2014 Mar;103(3):341-363. [doi: 10.1016/j.diabres.2013.10.012] [Medline: 24847517]
2. National Institute for Health and Care Excellence (UK). Outcomes and risks for woman and baby. In: Diabetes in Pregnancy: Management of Diabetes and Its Complications from Preconception to the Postnatal Period. London: National Collaborating Centre for Women's and Children's Health; 2015.

3. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008 May 08;358(19):1991-2002. [doi: 10.1056/NEJMoa0707943] [Medline: 18463375]
4. Harris DL, Weston PJ, Signal M, Chase JG, Harding JE. Dextrose gel for neonatal hypoglycaemia (the Sugar Babies Study): a randomised, double-blind, placebo-controlled trial. Lancet 2013 Dec 21;382(9910):2077-2083. [doi: 10.1016/S0140-6736(13)61645-1] [Medline: 24075361]
5. Bellamy L, Casas J, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
6. Page KA, Romero A, Buchanan TA, Xiang AH. Gestational diabetes mellitus, maternal obesity, and adiposity in offspring. J Pediatr 2014 Apr;164(4):807-810 [FREE Full text] [doi: 10.1016/j.jpeds.2013.11.063] [Medline: 24388326]
7. Ornoy A. Prenatal origin of obesity and their complications: Gestational diabetes, maternal overweight and the paradoxical effects of fetal growth restriction and macrosomia. Reprod Toxicol 2011 Sep;32(2):205-212. [doi: 10.1016/j.reprotox.2011.05.002] [Medline: 21620955]
8. Dabelea D, Hanson RL, Lindsay RS, Pettitt DJ, Imperatore G, Gabir MM, et al. Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. Diabetes 2000 Dec;49(12):2208-2211. [doi: 10.2337/diabetes.49.12.2208] [Medline: 11118027]
9. Rosenberg TJ, Garbers S, Lipkind H, Chiasson MA. Maternal obesity and diabetes as risk factors for adverse pregnancy outcomes: differences among 4 racial/ethnic groups. Am J Public Health 2005 Sep;95(9):1545-1551. [doi: 10.2105/AJPH.2005.065680] [Medline: 16118366]
10. American Diabetes Association. 14. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S183-S192. [doi: 10.2337/dc20-S014] [Medline: 31862757]
11. Martis R, Crowther CA, Shepherd E, Alsweiler J, Downie MR, Brown J. Treatments for women with gestational diabetes mellitus: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2018 Aug 14;8:CD012327 [FREE Full text] [doi: 10.1002/14651858.CD012327.pub2] [Medline: 30103263]
12. Nigam A, Sharma S, Varun N, Munjal YP, Prakash A. Comparative analysis of 2-week glycaemic profile of healthy versus mild gestational diabetic pregnant women using flash glucose monitoring system: an observational study. BJOG 2019 Aug;126 Suppl 4:27-33. [doi: 10.1111/1471-0528.15849] [Medline: 31257712]
13. Cosson E, Baz B, Gary F, Pharisien I, Nguyen MT, Sandre-Banon D, et al. Poor reliability and poor adherence to self-monitoring of blood glucose are common in women with gestational diabetes mellitus and may be associated with poor pregnancy outcomes. Diabetes Care 2017 Sep;40(9):1181-1186. [doi: 10.2337/dc17-0369] [Medline: 28724718]
14. Sousa AMDS, Fiuza D, Mikami FCF, Abrão KC, Francisco RPV, Zugaib M. Evaluation of information retention and adherence to treatment in patients with gestational diabetes mellitus after multidisciplinary group. Rev Assoc Med Bras (1992) 2016;62(3):212-217 [FREE Full text] [doi: 10.1590/1806-9282.62.03.212] [Medline: 27310543]
15. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
16. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-115 [FREE Full text] [doi: 10.1177/1932296814556506] [Medline: 25361643]
17. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: A Feasibility Study for a Randomized Controlled Trial Exploring the Potential for Telemedicine in the Diabetes Care of Those with Gestational Diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
18. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: Randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
19. Alqudah A, McMullan P, Todd A, O'Doherty C, McVey A, McConnell M, et al. Service evaluation of diabetes management during pregnancy in a regional maternity hospital: potential scope for increased self-management and remote patient monitoring through mHealth solutions. BMC Health Serv Res 2019 Sep 13;19(1):662 [FREE Full text] [doi: 10.1186/s12913-019-4471-9] [Medline: 31514743]
20. Al-Ofi EA, Mosli HH, Ghamri KA, Ghazali SM. Management of postprandial hyperglycaemia and weight gain in women with gestational diabetes mellitus using a novel telemonitoring system. J Int Med Res 2019 Feb;47(2):754-764 [FREE Full text] [doi: 10.1177/0300060518809872] [Medline: 30442052]
21. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
22. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]

23. Carral F, Ayala MDC, Fernández JJ, González C, Piñero A, García G, et al. Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. Diabetes Technol Ther 2015 May;17(5):349-354. [doi: 10.1089/dia.2014.0223] [Medline: 25629547]
24. International Association of Diabetes Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010 Mar;33(3):676-682 [FREE Full text] [doi: 10.2337/dc09-1848] [Medline: 20190296]
25. American Diabetes Association. 14. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S165-S172. [doi: 10.2337/dc19-S014] [Medline: 30559240]
26. Martis R, Brown J, Crowther CA. Views and Experiences of New Zealand Women with Gestational Diabetes in Achieving Glycaemic Control Targets: The Views Study. J Diabetes Res 2017;2017:2190812 [FREE Full text] [doi: 10.1155/2017/2190812] [Medline: 29226155]
27. Wah YYE, McGill M, Wong J, Ross GP, Harding A, Krass I. Self-management of gestational diabetes among Chinese migrants: A qualitative study. Women Birth 2019 Feb;32(1):e17-e23. [doi: 10.1016/j.wombi.2018.03.001] [Medline: 29693546]
28. Dennison RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on lifestyle changes to reduce the risk of developing Type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for practice. Diabet Med 2019 Jun;36(6):702-717 [FREE Full text] [doi: 10.1111/dme.13926] [Medline: 30723968]
29. Kolivand M, Rahimi MA, Keramat A, Shariati M, Emamian MH. Effect of a new self-care guide package on maternal and neonatal outcomes in gestational diabetes: A randomized control trial. J Diabetes 2019 Feb;11(2):139-147. [doi: 10.1111/1753-0407.12827] [Medline: 30022620]
30. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
31. Rasekaba TM, Furler J, Young D, Liew D, Gray K, Blackberry I, et al. Using technology to support care in gestational diabetes mellitus: Quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
32. Suh S, Kim JH. Glycemic variability: How do we measure it and why is it important? Diabetes Metab J 2015 Aug;39(4):273-282 [FREE Full text] [doi: 10.4093/dmj.2015.39.4.273] [Medline: 26301188]
33. Yang P, Lo W, He Z, Xiao X. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J Obstet Gynaecol Res 2018 Jul;44(7):1228-1234. [doi: 10.1111/jog.13669] [Medline: 29797375]
34. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: A systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
35. Lau Y, Htun TP, Wong SN, Tam WSW, Klainin-Yobas P. Efficacy of internet-based self-monitoring interventions on maternal and neonatal outcomes in perinatal diabetic women: A systematic review and meta-analysis. J Med Internet Res 2016 Aug 15;18(8):e220 [FREE Full text] [doi: 10.2196/jmir.6153] [Medline: 27526637]
36. Raman P, Shepherd E, Dowswell T, Middleton P, Crowther CA. Different methods and settings for glucose monitoring for gestational diabetes during pregnancy. Cochrane Database Syst Rev 2017 Oct 29;10:CD011069. [doi: 10.1002/14651858.CD011069.pub2] [Medline: 29081069]
37. Zilberman-Kravits D, Meyerstein N, Abu-Rabia Y, Wiznitzer A, Harman-Boehm I. The impact of a cultural lifestyle intervention on metabolic parameters after gestational diabetes mellitus: a randomized controlled trial. Matern Child Health J 2018 Jun;22(6):803-811. [doi: 10.1007/s10995-018-2450-0] [Medline: 29411251]
38. Koivusalo SB, Rönö K, Klemetti MM, Roine RP, Lindström J, Erkkola M, et al. Gestational Diabetes Mellitus Can Be Prevented by Lifestyle Intervention: The Finnish Gestational Diabetes Prevention Study (RADIEL): A Randomized Controlled Trial. Diabetes Care 2016 Jan;39(1):24-30. [doi: 10.2337/dc15-0511] [Medline: 26223239]

Abbreviations

BG: blood glucose
CGMS: continuous glucose monitoring system
GDM: gestational diabetes mellitus
OGTT: oral glucose tolerance test

Edited by L Buis; submitted 25.07.20; peer-reviewed by B Richardson, W Na; comments to author 04.11.20; revised version received 22.11.20; accepted 29.03.21; published 25.05.21.

Please cite as:
Tian Y, Zhang S, Huang F, Ma L
Comparing the Efficacies of Telemedicine and Standard Prenatal Care on Blood Glucose Control in Women With Gestational Diabetes Mellitus: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e22881
URL: https://mhealth.jmir.org/2021/5/e22881
doi:10.2196/22881
PMID:33783365

Original Paper

A Smartphone App (TRIANGLE) to Change Cardiometabolic Risk Behaviors in Women Following Gestational Diabetes Mellitus: Intervention Mapping Approach

Anne Lotte Potzel[1,2,3], Dr rer biol hum; Christina Gar[1,2,3], Dr rer biol hum; Jochen Seissler[1,2,3], MD; Andreas Lechner[1,2,3], MD

[1]Diabetes Research Group, Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany

[2]CCG Type 2 Diabetes, Helmholtz Zentrum München, Munich, Germany

[3]German Center for Diabetes Research, Neuherberg, Germany

Corresponding Author:
Andreas Lechner, MD
Diabetes Research Group
Medizinische Klinik und Poliklinik IV
Klinikum der Universität München
Ziemssenstr. 1
Munich, 80336
Germany
Phone: 49 894 400 521 85
Fax: 49 894 400 533 74
Email: andreas.lechner@med.uni-muenchen.de

Abstract

Background: Gestational diabetes mellitus (GDM) is the most common complication during pregnancy and is associated with an increased risk for the development of cardiometabolic diseases. Behavioral interventions can reduce this risk, but current solutions insufficiently address the requirements for such a program. The systematic development of a scalable mobile health (mHealth) promotion program for mothers during the first years post-GDM may contribute to solving this problem.

Objective: The aim of this project was to systematically plan and develop a theory- and evidence-based mHealth intervention to change cardiometabolic risk behaviors in women during the first 5 years post-GDM that meets women's expected standards of commercial health apps.

Methods: The intervention mapping steps 1 to 4 structured the systematic planning and development of the mHealth program described in this paper. Steps 1 and 2 led to a theory- and evidence-based logic model of change for cardiometabolic health. Based on this model, the prevention program was designed (step 3) and produced (step 4) in cooperation with industrial partners to ensure a high technological standard of the resulting smartphone app for the iPhone (Apple Inc). Step 4 included a user study with women during the first 5 years post-GDM once a beta version of the app ("TRIANGLE") was available. The user study comprised 2 test rounds of 1 week (n=5) and 4 weeks (n=6), respectively. The tests included validated questionnaires on user acceptance, user logs, and think-alouds with semistructured interviews.

Results: The novel TRIANGLE app is among the first self-paced smartphone apps for individual habit change in the 3 lifestyle areas of physical activity, nutrition, and psychosocial well-being. The 3 core features—a challenge system, human coaching, and a library—address 11 behavioral determinants with 39 behavior change methods to support lifestyle changes. Participants in the user study showed a high acceptance, high perceived quality, and high perceived impact of the TRIANGLE app on their health behaviors. Participants tested the app regularly, used it intuitively, and suggested improvements. We then adapted the TRIANGLE app according to the insights from the user study before the full TRIANGLE program production.

Conclusions: The intervention mapping approach was feasible to plan and develop an innovative and scalable smartphone solution for women during the first 5 years post-GDM. The resulting TRIANGLE intervention has the potential to support behavior change for cardiometabolic disease prevention. However, the app needs further refinement and testing in clinical trials. Intervention mapping steps 5 (implementation plan) and 6 (evaluation plan) may support the integration of the TRIANGLE intervention into routine care.

Trial Registration: German Clinical Trials Register DRKS00012736; https://www.drks.de/DRKS00012736

(JMIR Mhealth Uhealth 2021;9(5):e26163) doi:10.2196/26163

KEYWORDS

mHealth; diabetes prevention; health behavior; cardiometabolic disease; gestational diabetes mellitus; smartphone app; intervention mapping

Introduction

Background

Gestational diabetes mellitus (GDM) has become a public health problem as the most common complication during pregnancy, with rising numbers globally [1]. Following GDM, women have a 10-fold higher risk for type 2 diabetes [2] and a high risk for metabolic syndrome and other cardiometabolic diseases [3,4]. Previous lifestyle interventions among women with a history of GDM who had already developed impaired glucose tolerance reduced the risk of developing type 2 diabetes by approximately 50% [5]. However, recent research highlighted the importance of starting interventions earlier, particularly in the first year postpartum [6], as the years following GDM offer a window of opportunity for preventive lifestyle interventions to prevent the progression of cardiometabolic disturbances early on [7].

Prior Work and Proposed Solution

Qualitative research showed that lifestyle interventions within the 10 years following GDM face specific challenges such as prioritization of the family and a mother's perceived lack of resources [8]. These challenges are reflected in the modest effects of previous behavior interventions in the first few years post-GDM [6,7]. Hence, experts call for innovative, flexible, and personalized interventions specific to the needs of women during the first few years post-GDM and the socio-ecological contexts of health behaviors [9]. These needs include a practical and accessible tool for behavior change support that is compatible with daily family life [9-11]. Mobile health (mHealth) solutions offer easy access, use in daily family life, and integrated behavior change methods [12,13]. At present, the ideal mHealth platform for women post-GDM appears to be smartphone apps since they are frequently used for health purposes by women during the first few years post-GDM [14].

Aim of This Project

The aim of this project was to systematically plan and develop a theory- and evidence-based mHealth intervention for women during the first 5 years post-GDM that supports behavior change in several lifestyle areas. To ensure that the standards of both academia and industry were met, we decided to develop a smartphone app—the TRIANGLE app—with industrial partners using the intervention mapping protocol. Thereby, we aimed at both tackling some of the limitations from earlier prevention programs for women post-GDM [7,9] and testing the usability of the novel smartphone app with the intended users. This paper provides a detailed description of the systematic planning and development of the TRIANGLE intervention, including the user study.

Methods

Systematic Planning and Development of the TRIANGLE Intervention

We planned and developed the TRIANGLE intervention with the intervention mapping approach [15] as much as possible with the given resources. The process was guided by an understanding of the intervention context through the social-ecological model, systems thinking, stakeholder involvement, and ethical principles to identify and decide on the best intervention points. Further, we applied theory- and evidence-based practices using the 6 intervention mapping core processes as overarching principles. The core processes supported the use of different data sources for effective decision making by guiding us to pose a question related to the problem, brainstorm answers, review the literature, access and use theory, consider the need for new research, and define the working list of answers [15]. The 4 of the 6 intervention mapping steps that were applied in the scope of this project were a needs assessment with a logic model of the problem (step 1), a logic model of change (step 2), program design (step 3), and program production (step 4). These 4 steps provide the prerequisite for the 2 remaining steps—implementation planning (step 5) and evaluation planning (step 6)—not included in this publication.

TRIANGLE User Study

As part of intervention mapping step 4, the TRIANGLE user study (DRKS trial registration: DRKS00012736) tested the novel TRIANGLE app in 2 test rounds with 5 or more participants of the intended target group (women in their first 5 years post-GDM) in a mixed methods design. The user study was conducted at the Medical Center of the Ludwig Maximilian University of Munich. The content of the tested TRIANGLE app was limited for the purpose of this study (Multimedia Appendix 1). Participants gave their written informed consent and signed the TRIANGLE app's privacy statement. The study was approved by the ethics committee of the Medical Faculty of the Ludwig Maximilian University of Munich.

Study Participants

Participants were primarily recruited by phone from the patient base of the Medical Center of the Ludwig Maximilian University of Munich between mid-June and mid-July 2017. The inclusion criteria were a medical diagnosis of GDM in a recent pregnancy, 3 months to 5 years postpartum, postnatal core muscle recovery (assessed by enquiring about the completion of a 10-week course in the German health care system and a question on postnatal sporting restrictions), ownership of an iPhone 5 to 7 Plus (Apple Inc), and fluency in German. The exclusion criteria comprised age <18 years, current pregnancy, cardiopulmonary disease or restrictions in the locomotor system contraindicating a sports intervention, gastrointestinal disease contraindicating a nutrition intervention, psychiatric disease requiring therapy, other serious

illness contraindicating a lifestyle intervention according to the principal investigator, and alcohol or drug abuse.

Data Collection and Handling

Data collection comprised 2 visits (visit 1 [V1] and visit 2 [V2]) at the Medical Center of the Ludwig Maximilian University of Munich (details in Multimedia Appendix 2) with questionnaires, think-alouds with semistructured interviews (test setup in Multimedia Appendix 3), and app user logs collected during and between the 2 visits (Multimedia Appendix 4). Participants used the TRIANGLE app for a predefined period between the 2 study visits: 1 week (+ maximum 7 days) in group 1 and 4 weeks (± maximum 7 days) in group 2.

At V1, participants installed the TRIANGLE app on their iPhone and accessed their personalized account with an individual login code. They further received a Garmin vívosmart HR fitness tracker (Garmin Ltd), a step stool, and the TRIANGLE paper notepad. Participants were instructed to use the app on 5 or more days/week for the duration of the study, to conduct the fitness self-test, and to use the additional program materials. The client-server system of the TRIANGLE intervention tracked their activities on the app (Multimedia Appendix 4). The questionnaires completed during V1 included a social anamnesis, medical history, smartphone and app usage, the System Usability Scale (SUS) [16], and the user version of the Mobile Application Rating Scale (uMARS) [17]. The SUS and uMARS were also assessed during V2.

During a one-on-one think-aloud session followed by a semistructured interview, participants first received a standardized introduction video to get familiar with the test method (ie, to think aloud) before completing 30 small tasks from a standardized test protocol in the TRIANGLE app. An investigator (ALP) read aloud the tasks one by one and both observed and recorded (using audio and the iPhone screen, upon a user's consent only) how a user installed and registered on the TRIANGLE app, attended the introduction video, and navigated through the features. Thereby, the investigator was seated outside the user's field of vision (Multimedia Appendix 3) and noted the user's actions in the TRIANGLE app, expressed thoughts, questions, problems, and body language in a respective checklist structured per task. If a user interrupted his or her thinking aloud, the investigator gave prompts such as, "What is on your mind?" The semistructured interviews (audio-recorded upon the user's consent only) addressed the user's overall impression and subjective experience with the TRIANGLE app, as well as some specifics (at V2 only) about the didactic design, content, expectations, and initial habit change. We aimed at timely insights and improvements via an explorative rapid thematic analysis before the start of the clinical pilot study. Hence, the investigator reviewed the screen and audio recordings and amended the written notes in case of a missed observation during the session instead of preparing a full transcript of each record.

Collected data were pseudonymized (4-digit code) and TRIANGLE app data were encrypted (2048-bit Secure Sockets Layer and end-to-end encryption) and stored on the server of the Medical Center of the Ludwig Maximilian University of Munich.

Statistics

Each TRIANGLE user group contained 5 or more participants to detect approximately 80% of the issues during usability testing [18,19]. Because of the small group sizes, no statistical group comparisons were conducted. We analyzed all data in Microsoft Excel 2016 MSO (Microsoft Corporation) and Tableau Desktop 2019.3 (Tableau Software, LLC). Values are presented as counts (n) with percentages or as mean (SD). The analyses were based on all participants with the respective complete outcome measure. We excluded app activities collected during the think-aloud sessions for the app usage analyses between the 2 visits. Data from the observations during the think-alouds and the semistructured interview questions were analyzed in an explorative rapid thematic analysis with the respective numbers of participants [20,21].

Results

Systematic Planning and Development of the TRIANGLE Intervention

Step 1: Needs Assessment and Logic Model of the Problem

Task 1.1: Establish and Work With a Planning Group

The core team of the project (ALP and AL) is part of the Diabetes Research Group in the Medical Center of the Ludwig Maximilian University of Munich and headed the expert work groups. Expert work groups and industrial contractors covered the areas of software engineering/user interface design (n=5), sports science/personal training (n=4), nutritional science/counselling (n=3), psychological therapy/behavioral coaching (n=3), and multimedia content creation (n=2). The Diabetes Research Group is a clinical cooperation group of the Helmholtz Zentrum Muenchen and has run the 10-year Prediction, Prevention and Sub-classification of type 2 Diabetes (PPSDiab) observational study with women post-GDM since 2012 [22]. Hence, the study team gained important insights on the target group using, for example, cardiometabolic exercise testing and validated lifestyle questionnaires [23-25].

Task 1.2: Conduct a Needs Assessment to Create a Logic Model of the Problem (Multimedia Appendix 5)

Based on expertise and the literature [6,26,27], we defined the target group as women of reproductive age (predominantly aged 18-45 years), with ≥1 child in the household, ≥1 recent pregnancy complicated by GDM, in the extended postpartum period (maximum 5 years after delivery), and at high risk for or with cardiometabolic disturbances (as indicated by a history of GDM and the following specifications). We identified 2 main subgroups: women who were overweight with a BMI ≥23 kg/m^2 and women with a BMI <23 kg/m^2. We chose the threshold of BMI ≥23 kg/m^2 previously applied to Asian populations [28] instead of the traditional threshold of BMI ≥25 kg/m^2 [29] to define overweight since the women in the target group were younger [30] than traditional at-risk cohorts [29,31]. Therefore, weight reduction should be attempted already at a BMI ≥23 kg/m^2. We further amended the health problem of type 2

diabetes by the following 4 main clusters of related cardiometabolic risk factors and/or disturbances to identify related risk behaviors: overweight/obesity [32,33], high blood pressure [3,34], dysglycemia [4,35], dyslipidemia [36,37], and combinations thereof [4,38]. Our related search for quality-of-life implications pointed toward little or no immediate impact of GDM on quality of life after delivery [39] (apart from some vulnerable subgroups, such as those with postpartum depression [40] or obesity [41]) versus long-term impairments in all major quality-of-life domains [42] in the presence of type 2 diabetes and/or related cardiometabolic disturbances [43,44], starting with a prediabetic state [45].

Next, we categorized the 20 main cardiometabolic risk behaviors (Textbox 1 [32,46-70]) into the lifestyle areas of physical activity, nutrition, psychosocial well-being (including sleep), and "other."

Textbox 1. Cardiometabolic risk behaviors.

Nutrition habits

- Excess intake of energy-dense meals or snacks [46,47]
- Excess intake of animal-derived products, especially those high in total fat/saturated fat [48]
- Excess intake of (ultra-)processed food [47]
- Excess total energy intake [32]
- Excess intake of caloric drinks such as soda [49] and alcohol [50]
- Insufficient intake of unprocessed to minimally processed food [51]
- Insufficient intake of (fresh) plant products [52]
- Insufficient intake of water, plain tea, or coffee [53,54]
- Nonadherence to recommended healthy dietary patterns and macronutrient quality [55-57]
- Insufficient control of eating behavior or emotional eating [58]

Physical activity habits

- Insufficient daily physical activity or sedentary behavior [59]
- Insufficient exercise or exercise intensity [60]

Psychological and sleep habits

- Pessimistic thinking style [61]
- Insufficient control of negative emotions [8,62]
- Poor stress management or problem solving [8,63]
- Insufficient meditation, positive emotion, optimistic thinking, or mind-body exercises [64,65]
- Insufficient enjoyable leisure time activities [66]
- Insufficient sleep-enhancing behaviors [67,68]

Other behaviors

- No or limited breastfeeding [69]
- Smoking [70]

The most common behavioral theories used in intervention mapping pointed to key theoretical constructs for personal determinants of behavior [15] that we used to structure the literature search for our priority population. Relevant theories included theories on automatic behavior and habits [71,72], goal setting [73], information processing/persuasive communication [74,75], process models of behavior change [76-78], and social cognitive models [79,80]. Both the behavioral theories and previous qualitative research indicated 11 main personal determinants for cardiometabolic risk behaviors in women post-GDM: habit [81], commitment [82], behavioral knowledge [83], perceived risk [84], perceived barriers to behavior [85], perceived behavioral skills [86], perceived self-efficacy [84], outcome expectations [87]/attitudes [8,81], perceived social norms [83], self-image [8], and emotional reaction to behavior [81].

Task 1.3: Describe the Context for the Intervention, Including the Population, Setting, and Community

The planned intervention setting was daily family life of a mother during the first 5 years post-GDM, with high demands for intervention flexibility [8] (eg, due to irregular days, financial restrictions, and/or travel during parental leave). The intervention further focused on an active relationship between health care practitioners and a woman in the first few years post-GDM to harness the momentum of the GDM diagnosis [82] and the need for personal support [88] in this locally

scattered niche population. The first version of the intervention targeted behavior changes in the German cultural setting, given a high smartphone usage among mothers during the first years postpartum and numerous physical activity options.

Task 1.4: State Program Goals

The aim of the TRIANGLE intervention is to see a decrease in the incidence of type 2 diabetes following GDM in Germany by 30% at the 6-year follow-up after the intervention.

Step 2: Logic Model of Change

Task 2.1: State Expected Outcomes for Behavior

Similar to the Mothers After Gestational Diabetes in Australia Diabetes Prevention Program (MAGDA-DPP) trial [89], we adapted the 5 traditional behavioral end points as pursued by diabetes prevention programs (DPPs [29,31]) to women during the first 5 years post-GDM and added 2 outcomes for improved psychosocial well-being (including sleep) [58] and intervention adherence [90] (Textbox 2).

Textbox 2. Expected behavioral outcomes for cardiometabolic disease prevention during the first 5 years after gestational diabetes mellitus.

- Physical activity of moderate to high intensity for ≥150 minutes/week
- Dietary fiber intake ≥15 g/1000 kcals
- Percentage of fat intake <30% of total energy intake
- Percentage of saturated fatty acid intake <10% of total energy intake
- Body weight reduction of ≥5% if BMI is ≥23 kg/m^2; body weight maintenance if BMI is <23 kg/m^2
- Increased psychosocial well-being and sleep, and decreased stress perception
- Intervention adherence and enhanced self-management

Task 2.2: Specify Performance Objectives for Behavioral Outcomes

We delineated each of the 7 expected behavioral outcomes (Textbox 2) into concrete actions that women in the first 5 years post-GDM need to perform to achieve a specific outcome. Thereby, we distinguished between preparatory one-time actions (eg, “Remove ultraprocessed food products from your home food supplies”) and habitual actions (eg, “Walk at least 10,000 steps” [daily] or “Plan your family meals for the next week” [weekly]). Our resulting list of 81 performance objectives considered the family context, flexible settings in daily life, and different entry levels with logical sequences for skill building (Textbox 3 and Multimedia Appendix 6). The high number of performance objectives allowed for tailoring of the 2 subgroups (BMI ≥23 kg/m^2 vs BMI <23 kg/m^2) and further individualization. Aiming primarily at long-term habit change, we excluded some of the identified cardiometabolic risk behaviors (eg, temporary behaviors, such as breastfeeding).

Textbox 3. Exemplary performance objectives (POs) for the first behavioral outcome: physical activity of moderate to high intensity for ≥150 minutes/week.

- PO.1.1: Use a fitness tracker for daily step count.
- PO.1.2: Monitor the average daily step count of 1 week.
- PO.1.3: If walking less than 10,000 steps a day, decide to gradually increase daily steps to at least 10,000.
- PO.1.4: Depending on the personal starting point, walk at least 5500 steps on 3 days a week, then every day; walk at least 8000 steps on 3 days a week, then every day; and, ultimately, walk at least 10,000 steps on 3 days a week, then every day.
- PO.1.5: Disrupt longer sedentary periods every 30 minutes.
- PO.1.6: If not currently exercising, decide to initiate an exercise routine.
- PO.1.7: Conduct the fitness self-test at home.
- PO.1.8: If current exercise volume is less than 150 minutes of moderate to high intensity per week, decide to gradually increase exercise volume to reach the recommended level.
- PO.1.9: Plan own exercise regimen with health care practitioner including frequency, intensity, time, type, volume, and progression of exercise, with specific, measurable, action-oriented, realistic, timely, and self-determined goals.
- PO.1.10: Implement own exercise regimen.
- PO.1.11: Use fitness tracker for heart rate monitoring during exercise units.
- PO.1.12: Engage in active regeneration on nonexercise days.
- PO.1.13: Engage in active transportation.

Task 2.3: Select Determinants of Behavioral Outcomes

For optimal intervention outcomes, we formed 7 clusters of the 11 interrelated theory- and evidence-based behavioral determinants for women during the first 5 years post-GDM (Textbox 4).

Textbox 4. Clustered personal determinants of the performance objectives.

- Perceived risk, behavioral knowledge, and commitment
- Perceived barriers
- Perceived skills and self-efficacy
- Outcome expectations and attitudes
- Perceived social norms
- Self-image and habits
- Emotional reaction to behavior

Task 2.4: Construct Matrices of Change Objectives

We crossed every performance objective with suitable determinants to uncover necessary changes for women during the first 5 years post-GDM (see example in Table 1).

Table 1. Change objectives (COs) for performance objective 1.12: Engage in active regeneration on nonexercise days.

Clustered personal determinants	COs
Perceived risk, behavioral knowledge, and commitment	• CO.1.12.1: Get informed about the benefits of active regeneration and learn strategies for how to do it. • CO.1.12.2: Acknowledge the habitual character of regeneration and the need to change environmental cues to engage in active regeneration. • CO.1.12.3: Decide to engage in active regeneration on nonexercise days.
Perceived barriers	• CO.1.12.4: Get informed about possible perceived barriers to engaging in active regeneration and identify personal barriers. • CO.1.12.5: Get informed about possible solutions to overcome perceived barriers to engaging in active regeneration and implement the most suitable solutions. • CO.1.12.6: Expect and resist hindering social pressure by family members or the wider social network when engaging in active regeneration.
Perceived skills and self-efficacy	• CO.1.12.7: Express confidence in ability to engage in active regeneration or to learn how to do so. • CO.1.12.8: Feel capable of noticing automaticity on regeneration days and altering cues that trigger engaging in active regeneration.
Outcome expectations and attitudes	• CO.1.12.9: Expect that engaging in active regeneration leads to less aching muscles and cardiometabolic health benefits. • CO.1.12.10: Feel positive about engaging in active regeneration.
Perceived social norms	• CO.1.12.11: Notice that most physically fit individuals consistently engage in active regeneration and find role models in their own social network. • CO.1.12.12: Notice that engaging in active regeneration does not need approval by others.
Self-image and habits	• CO.1.12.13: Consistently maintain an active regeneration routine until habitual. • CO.1.12.14: Identify as a healthy homemaker and role model who guides own children and partner to enjoy being active together.
Emotional reaction to behavior	• CO.1.12.15: Expect initial discomfort when not used to engaging in active regeneration. • CO.1.12.16: Notice that engaging in active regeneration is fun and does not translate to constraints. • CO.1.12.17: Feel great about having engaged in active regeneration.

Task 2.5: Create a Logic Model of Change (Multimedia Appendix 7)

The logic model of the problem (intervention mapping step 1; Multimedia Appendix 5) and the matrices of change objectives (Task 2.4; Table 1) informed our logic model of behavior change for cardiometabolic health during the first 5 years post-GDM.

Step 3: Program Design

Task 3.1: Generate Program Themes, Components, Scope, and Sequence

We collected program requirements for women during the first few years post-GDM as assessed in previous qualitative studies via interviews, observations, open-ended surveys, and focus

groups [8]. We then translated the requirements into the following 4 clusters of mHealth specifications: (1) 3 lifestyle areas under the umbrella of behavioral psychology, (2) family context, (3) individual content and participation, and (4) other specific needs of women during the first years post-GDM. We stressed the intertwined nature of the 3 main lifestyle areas that form the intervention modules (physical activity, nutrition, and psychosocial well-being) by using a symbolic health triangle as the logo and naming the smartphone app TRIANGLE. The branding of the TRIANGLE intervention with a challenge system reflects the challenging nature of multiple habit changes and the overarching module of behavior change. We planned to deliver all modules digitally (Multimedia Appendix 8) with a flexible dosage and sequence to account for individual differences in habit formation. Essential program components include multimedia knowledge transfer, individual behavior change support, and skill training. Additional program materials comprise a fitness wrist band, a step stool, and the TRIANGLE paper notepad.

Tasks 3.2 and 3.3: Choose and Translate Theory- and Evidence-Based Change Methods Into Practical Applications

In line with the intervention mapping–related taxonomy of behavior change methods [91], we selected 39 behavior change methods (Table 2) to address all specified change objectives and translated them into app features. Thereby, we focused on habit formation as shared characteristics of the 3 lifestyle areas. This resulted in the following 3 core features of the app: (1) an interactive challenge system, (2) individual human coaching, and (3) a library. An associated coaching platform mirrored these features for content management and individualization. All interactions were designed to be short in duration; for example, a participant may accept a challenge, the personal coach then motivates the participant to commit to this challenge despite the delayed health benefits, and the library provides a short article on the benefits of early commitment.

Table 2. Behavior change methods to address clustered personal determinants.

Clustered personal determinants	Behavior change methods
Perceived risk, behavioral knowledge, and commitment	• Tailoring • Chunking • Advance organizers • Using imagery • Discussion • Framing • Environmental reevaluation • Credible information • Individualization • Participation • Technical assistance • Persuasive communication • Consciousness raising
Perceived barriers	• Participatory problem solving • Planning coping responses • Resistance to social pressure
Perceived skills and self-efficacy	• Guided practice • Enactive mastery experiences • Verbal persuasion • Improving emotional states • Self-monitoring • Technical assistance • Goal setting • Setting graded tasks • Cue altering
Outcome expectations and attitudes	• Arguments • Direct experience • Belief selection • Providing contingent rewards • Elaboration
Perceived social norms	• Information about others' approval • Resistance to social pressure • Mobilizing social support
Self-image and habits	• Counterconditioning • Implementation intentions • Stimulus control • Early commitment • Public commitment • Technical assistance • Reinforcement • Self-reevaluation • Environmental reevaluation
Emotional reaction to behavior	• Self-reevaluation • Improving physical and emotional states • Direct experience • Feedback

Step 4: Program Production

Task 4.1: Refine Program Structure and Organization

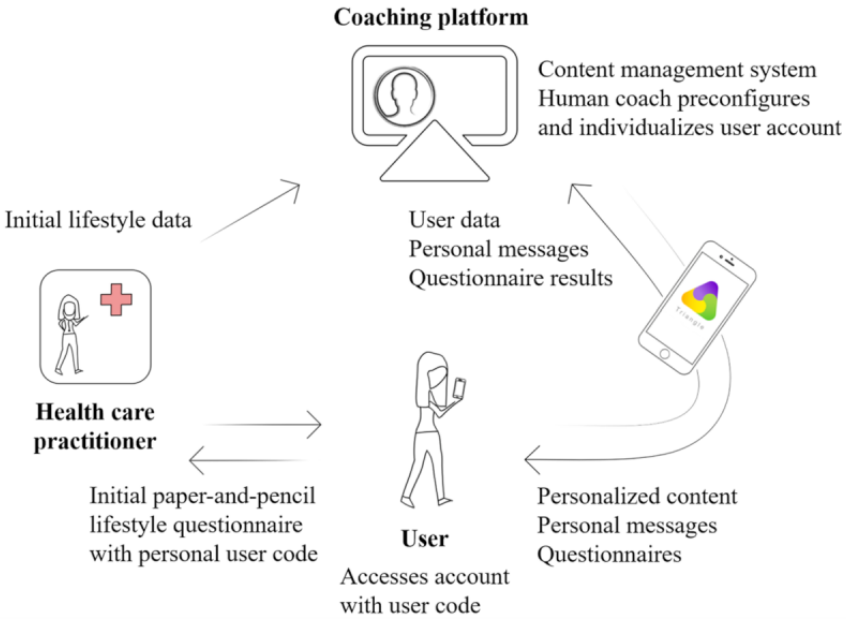
Our market analysis showed a gap between the identified requirements and current health promotion programs for women during the first 5 years post-GDM. Therefore, we decided to produce new program materials to address all identified requirements. Iterative processes helped us define the intervention modules (Multimedia Appendix 8) and the software requirements for the subfeatures of the app and the coaching platform. We prioritized the production of high-quality multimedia content as follows: 136 challenges and 133 library articles with an introduction and a medical background video, 54 videos related to physical activity (27 guided practices, 5 fitness self-tests, and 22 short tutorials), 8 psychological audio exercises, and 241 images. The challenges were mainly derived from the list of performance objectives (step 2) and contained options for tailoring and individualization. An initial

paper-and-pencil questionnaire and in-app questionnaires using the Apple ResearchKit informed the individualization process (eg, according to personal preferences and goals).

Task 4.2: Prepare Plans for Program Materials

The content of the TRIANGLE intervention needed to match the features of the planned app. Therefore, we developed design documents and style templates for the planned content (questionnaires, challenges, and library articles) and a scheme of the work and data flow between health care practitioners, the coaching platform, and app users/the app (Figure 1). Thereby, we considered smartphone-specific issues such as users' short attention span, screen limitations, and data protection. We secured a sociocultural fit of the intervention specific to Central European women during the first few years post-GDM in the multimedia content by involving female experts of childbearing age in all work groups. Further, we briefed all lifestyle work groups on the style templates to guarantee coherence.

Figure 1. TRIANGLE app onboarding and workflow.

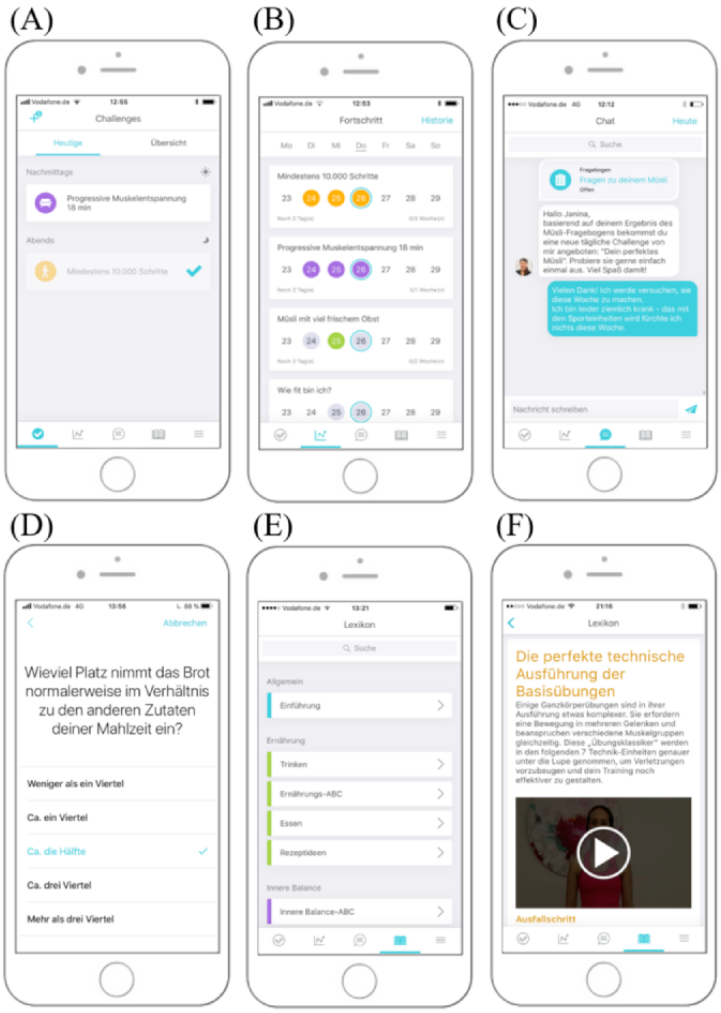


Task 4.3: Draft Messages, Materials, and Protocols

The software programming started with mock-ups in parallel with the content development by the lifestyle work groups. Given time and budget limitations, we decided to program the first version of the app for the iOS system for iPhones only, in line with the iOS developer human interface guidelines. The TRIANGLE app's core features were realized as follows (Figure 2): the interactive and individualized challenge system contains an activity screen, self-monitoring with progress visualization, and notifications/reminders. The individual human coaching is based on positive psychology, personal text messaging with a health care practitioner, in-app questionnaires, and personalized content/feedback. The library includes different modules with multimedia content and a keyword search.

We tested the app's features after each programming cycle, with thorough alpha testing of the app before the start of the user study. We specified which type of user data would be collected to allow for tailoring and individualization while complying with the General Data Protection Regulation. Further, we implemented encryption and authentication measures to protect against data misuse by third parties. A step-by-step guide supported the registration process of a participant. Our chosen lifestyle color scheme is yellow for physical activity, green for nutrition, and purple for psychosocial well-being, while the overarching color scheme is light blue and light grey, as often used in health care [58,92,93].

Figure 2. Screenshots of the TRIANGLE app's core features. Challenge system: (A) activity screen based on active challenges and (B) weekly progress visualization per challenge. Coaching: (C) chat with personal coach and (D) in-app questionnaires. Library: (E) categories per theme and (F) exemplary article.



Task 4.4: Pretest, Refine, and Produce Materials

The pretest was conducted in the scope of the TRIANGLE user study.

TRIANGLE User Study

The demographic characteristics of the user sample are shown in Table 3.

Table 3. Demographic characteristics of participants in the TRIANGLE user study.

Characteristics	Total (n=11)	Group 1 (n=5)	Group 2 (n=6)
Age (years), mean (SD)	36.6 (2.2)	35.4 (1.5)	37.7 (2.3)
Partner status: with partner, n (%)	10 (90.9)	5 (100)	5 (83.3)
Number of children, n (%)			
1	5 (45.5)	3 (60.0)	2 (33.3)
2	6 (54.5)	2 (40.0)	4 (66.7)
Work status			
Not working or on maternity leave, n (%)	2 (18.2)	1 (20.0)	1 (16.7)
Working, n (%)	9 (81.8)	4 (80.0)	5 (83.3)
Work hours, mean (SD)	27.4 (6.7)	24.5 (5.3)	29.8 (7.2)
Language level: native speaker, n (%)	9 (81.8)	5 (100)	4 (66.7)
Education, n (%)			
High school diploma	3 (27.3)	0 (0)	3 (50.0)
College or university degree	8 (72.7)	5 (100)	3 (50.0)
Treatment for gestational diabetes mellitus: insulin, n (%)	7 (63.6)	2 (40.0)	5 (83.3)
iPhone usage: >10 times per day, n (%)	9 (81.8)	4 (80.0)	5 (83.3)
Health app usage, n (%)			
Current	2 (18.2)	2 (40.0)	0 (0)
Previous	5 (45.5)	3 (60.0)	2 (33.3)
Never	4 (36.4)	0 (0)	4 (66.7)
Gadget usage, n (%)			
Current	1 (9.1)	1 (20.0)	0 (0)
Previous	1 (9.1)	0 (0)	1 (16.7)
Never	9 (81.8)	4 (80.0)	5 (83.3)

Process Measures of the Intervention

Group 1 (n=5) tested the TRIANGLE app for approximately 1 week (9.0 days, SD 2.3 days), while group 2 (n=6) tested it for approximately 4 weeks (27.3 days, SD 3.7 days). On the SUS, participants rated the app with a mean score of 87.3 (SD 8.8) points and 87.5 (SD 10.6) out of 100 points at V1 and V2, respectively (n=11; adjective rating: “excellent”). On the uMARS, participants gave an app quality mean score of 4.1 (SD 0.7) points and 4.2 (SD 0.8) points out of a total of 5.0 points at V1 and V2, respectively (n=8; n=3 missing; Figure 3).

The highest of the 4 uMARS objective quality subscores for the TRIANGLE app was the information mean score of 4.5 (SD 0.7) points (at both V1 and V2), while the lowest subscore was the engagement mean score of 3.9 (SD 0.9) points at V1 and 4.0 (SD 0.8) points at V2 (Figure 3). The app subjective quality mean score increased from 3.8 (SD 1.1) points at V1 to 4.3 (SD 0.9) points at V2. Similarly, the perceived impact mean score increased from 3.9 (SD 1.0) points at V1 to 4.4 (SD 0.9) points at V2.

Most participants tested the TRIANGLE app on a regular basis, some almost daily (Figure 4A). App activity in the given test period varied by participant (Figure 4A), yet each participant tested each lifestyle area (Figure 4B). The app was used at almost any time of the day, especially in the evening, with a nocturnal pause from midnight to 4 AM (Figure 4C). Coaching times were restricted to between 7 AM and 4 PM on weekdays. Participants primarily used the challenge system by ticking off challenges and opening challenge descriptions (Figure 4D). These activities were followed by opened library articles and played guided practices. The coach showed a low activity in comparison with participants and primarily sent text messages.

Figure 3. Results of the user version of the Mobile Application Rating Scale in the TRIANGLE user study stratified by visit and subscale (n=8; all values as mean [SD]).

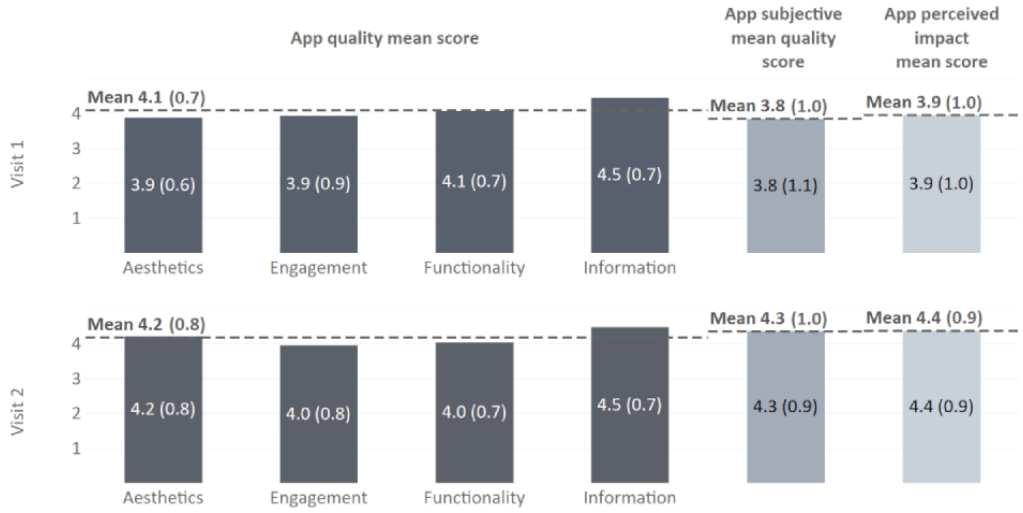
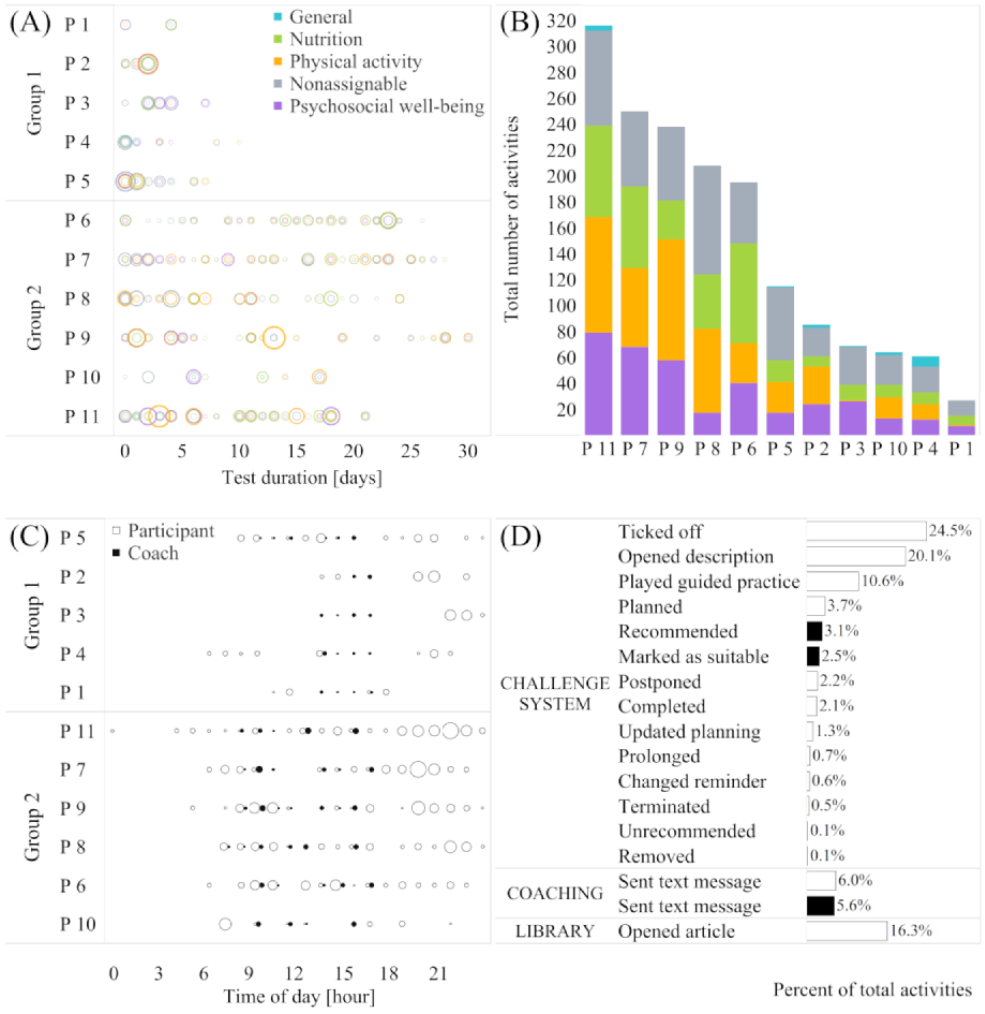


Figure 4. App usage in the TRIANGLE user study (n=11; group 1: n=5; group 2: n=6). (A) Number of app activities per participant over time stratified by theme; 1 circle per active day, with circle size reflecting the number of activities. (B) Total number of app activities per participant stratified by theme. (C) Number of app activities per participant and time of day; 1 circle per active hour, with circle size reflecting the number of activities. (D) Percent of app activities per subfeature. P: participant.



Qualitative Measures of the Intervention

Overall, participants gave positive feedback on the TRIANGLE app during the think-alouds with semistructured interviews. Participants used the app intuitively, except for 1 participant who showed initial problems with the swiping gesture. Further, participants considered the 3 core features easy to use and helpful. The combined data from the think-alouds and semistructured interviews showed 7 themes in our exploratory analysis—onboarding process (Table 4), navigation bar, challenge system, coaching, library, settings, and other feedback—with positive and negative statements for each theme. All positive statements supported the current format of the intervention.

Table 4. Exemplary analysis of the think-alouds with semistructured interviews for the onboarding process in the TRIANGLE user study.

Onboarding subthemes and statements from participants	Number of participants (N=11)
Introductory video overall	
Positive statements	
Acceptable length	9
Good overall	1
Answers all relevant questions	1
Good explanation of the features	1
Negative statements	
Static image at the beginning is confusing	6
Nonprofessional speaker	1
Male voice for female app	1
Marginally lengthy	1
Too little animation	1
Should be available at any time	1
Introductory video content	
Positive statements	
Contains everything important	3
Informative value	2
Stresses the essential	2
Very good content	1
Acceptable content	1
Negative statements	
Missing explanation of successful week algorithm	1
Missing specifications of fitness tracking	1
Marginally dense content for length	1
Too technical	1
Requires multiple viewings	1
Intervention procedure not entirely clear	1
Missing specification of coaching response time	1
Introductory video layout	
Positive statements	
Clear overview	1
Nice layout	1
Good	1
Negative statements	
Navigation not clear enough	2
Physical fitness color not “yellow,” rather “orange”	1
Color themes not explained in right order	1
Chat screen as start page after the video	
Positive statements	

Onboarding subthemes and statements from participants	Number of participants (N=11)
Nice welcome	5
Looks familiar	1
Positively surprised	1
Personal	1
Pleasantly simple	1
Negative statements	
Not suitable	2
Confusing	2
First impression app	
Positive statements	
Appears like regular app	1
Nice overall layout	1
Pleasantly simple	1
Negative statements	
Empty home screen before adding challenges is confusing	1

To improve the TRIANGLE app, we phrased crucial negative statements as problem descriptions and formulated the resulting changes (Table 5). We prioritized refinements and revisions of the TRIANGLE intervention based on the frequency and relevance of user feedback and the feasibility within this funding period. The main features of the TRIANGLE app remained the same before and after the pretest (Figure 2, post-testing). Additional changes to the TRIANGLE app besides those listed in Table 5 included a revised introductory video; a recipe section; more specific content for the late postpartum phase, such as pelvic floor training; an information section about the app, colored headings and subheadings per lifestyle area; and a revised paper-and-pencil questionnaire. Based on these adaptations, we revised and produced all of the TRIANGLE program materials.

Table 5. Exemplary problem descriptions and resulting changes to the TRIANGLE app after the TRIANGLE user study.

Problem descriptions	Resulting changes
Navigation bar	
Calendar icon raised user expectations of a calendar with similar interactive subfeatures to the iPhone calendar	Progress graph icon for the progress visualization of challenges
Closed-book icon raised user expectations of an interactive note pad	Open-book icon for the library
Challenge system	
Users did not immediately notice newly available challenges	Colored number of new challenges
Users were confused by the diverging order of challenges between the home screen and the weekly challenge overview	Synchronized challenge order of home screen and weekly challenge overview
Users viewed the “postpone challenge” button as unnecessary given the “terminate challenge” button	Removed “postpone challenge” button
Users considered the location of the challenge buttons at the end of a challenge description as impractical and complicating access to action	Restructured order of text and buttons in a challenge description, with buttons located at the beginning of an accepted challenge
Users thought the challenge text with a question-answer structure was marginally lengthy	Reduced challenge text with a simplified chunking and advance organizer structure
Users indicated an overuse of structuring elements such as capital letters, bold letters, and italics	Minimal use of structuring elements
Challenge-specific: (1) Challenge 1.4.1: “Walk at least 3000 steps” did not apply to any user; (2) Challenge 6.1: “8 minutes of mindfulness” was considered too long by users; (3) Challenge 6.5.1: “Keep a sleep diary“ followed by challenge 6.5.2: “Analyze your sleep diary” meant too much work for users	Removed challenges 1.4.1, 6.1, 6.5.1, and 6.5.2 from the list of challenges
Coaching	
Users did not immediately notice a new coaching activity	Colored number of new coaching activities in the navigation bar
Some users were unmotivated without a sufficient number of motivational messages	Expanded database of motivational texts, some specific to each lifestyle area
The difference between the “cancel” button in a questionnaire and the “cancel” button for exiting the canceling procedure was not clear	Renamed second “cancel” button in questionnaire tool
Users were confused by the automatic transfer of questionnaire results to the coach upon completion without further notice	Added button for actively sending questionnaire results to the personal coach and a notification for successful transfer

Discussion

Principal Findings

The goal of this project was to plan and develop one of the first smartphone app–based interventions to change cardiometabolic risk behaviors in women during the first 5 years post-GDM [14,28,92,94]. The resulting TRIANGLE intervention confirmed the value of using the systematic intervention mapping framework to plan a complex health promotion program [95]. The iterative development process led to a detailed theory- and evidence-based intervention model that was translated into the TRIANGLE app and then tested by its intended users.

Systematic Planning and Development of the TRIANGLE Intervention

We attempted to tackle some of the limitations of previous interventions during the first few years post-GDM. One of these limitations was the use of no or a single cognitive theory and few behavior change methods [96]. The more behavior change methods a health promotion program applies, the more likely it will affect behavior [97]. Hence, we used a multitheory approach at the habit-goal interface, applying 39 behavior change methods. The applied behavior change methods addressed 11 behavioral determinants and complied with the respective parameters of each method. The TRIANGLE intervention is the first intervention to prioritize habit change for women during the first 5 years post-GDM based on evidence that most health behaviors are habitual [71] and women are susceptible to unhealthy habits during the first few years post-GDM [98].

The TRIANGLE app contains a comprehensive multimedia database of newly developed lifestyle content with evidence-based information and exercises at the habit level to meet the needs of women during the first 5 years post-GDM. The content includes psychosocial well-being along with the more common DPP goals for nutrition and physical activity [7]. This addition was suggested in recent reviews [58,99] and is neededeg to prevent postnatal depression in women post-GDM [99]. Compared with previous programs during the first 5 years post-GDM, our extensive list of 81 performance objectives distinguishes between preparatory and habitual actions. Further, our performance objectives include many subbehaviors and supportive behaviors to achieve a greater behavioral outcome.

In addition, we target major health decisions and communication goals as necessary actions. Few other programs during the first few years post-GDM address postpartum topics such as pelvic floor health and sleep disruptions [98], which are included in the TRIANGLE intervention.

In addition, we considered the real-life intervention context and created a practical tool for both health care practitioners and women during the first few years post-GDM [100]. This tool has the potential for widespread implementation, low-cost intervention delivery in daily life, and human coaching by multiple health care professionals. The resulting tool also shows the importance of academia-industry cooperation in order to fulfil both standards in mHealth, as recently requested by women in the first few years post-GDM [14]. Our academia-industry cooperation led to a highly usable and accepted mHealth product with unique themes, components, scope, and sequence, while most other programs during the first 5 years post-GDM were based on the DPP model. We set up strong technical assistance to support the program engagement goal of at least 5 out of 7 days per week versus the more traditional weekly or monthly interactions [101]. Further, we translated the components of skill training, behavior change support, and knowledge transfer into a unique challenge-coaching-library feature system with flexible pacing. In contrast to the flexible pacing desired by women in the first 2 years post-GDM [11], most other programs followed fixed sequences such as the DPP curriculum [89,102]. Current programs are testing different features for smartphone-based interventions in the first year post-GDM, varying degrees of automation, and cross-links to other platforms and thus will soon provide further insights [14,28,94].

Lastly, the systematic planning and development led to a transparent description of the underlying theory of the TRIANGLE intervention to inform future intervention models, related studies, or projects. To date, such a detailed description of a behavior intervention post-GDM is missing, making it difficult to compare and identify the most successful intervention components for this target group. Hence, our model may be of value for other work groups to replicate or learn from certain decisions at each intervention mapping step.

TRIANGLE User Study

The TRIANGLE user study demonstrated that all of the main components of the TRIANGLE intervention were accepted by the intended users (women during the first 5 years post-GDM), including a first beta version of the TRIANGLE app for the iPhone. The study confirmed the value of mixed methods usability testing [103]. The SUS ratings of participants indicated “excellent” system usability [104] of the TRIANGLE app. The uMARS quality ratings of the TRIANGLE app (ie, 4.2 out of 5.0 points at V2) lay just below the top range of uMARS quality ratings for health apps (ie, 4.3 points [105])—mainly because of the lower engagement mean score compared with the other subscores. This indicated a need to enhance engagement, yet may also be related to both the limited content and the limited coach interaction in the user study. The increase in the app perceived impact mean score after several days of app usage demonstrated that users perceived a higher influence of the app on their behavior over time, as they may have started to change their health behaviors.

Overall, a smartphone app seemed to match the needs of women during the first 5 years post-GDM because participants used the TRIANGLE app regularly at almost any time of the day throughout the different lifestyle modules. The regular overall usage of the TRIANGLE app confirmed the feasibility of our usage goal of 5 active days per week. However, just like for a similar app [93], usage varied widely between participants. Each participant used each lifestyle area and thus confirmed the value of the added psychosocial module, as suggested by a recent review [58]. The high usage of the interactive challenge system stressed the importance for women to check off completed tasks for self-monitoring [93] and to view their own progress [14] in the first years post-GDM. In contrast, coaching requirements were much lower and indicated a realistic scope for implementation in routine care, especially when considering that some coaching activities may be automated in the future.

During the think-aloud sessions, participants used the app intuitively and considered the features useful. Further, the positive user statements during the semistructured interviews confirmed the chosen features and lifestyle modules. We specified and clustered emerging usability issues and negative statements as problem statements to guide smaller adaptations before the full program production, similar to related apps [14,93]. The desire of single participants for new features will have to be quantified before changing the TRIANGLE app system. Some of these desired features for apps in the first year post-GDM are currently being tested by other work groups, such as a built-in pedometer to translate other types of physical activity into step counts [28].

Limitations

The development of the TRIANGLE intervention using the intervention mapping approach had several limitations. The comprehensive intervention mapping approach is resource consuming, and this project offered limited human and financial resources and a tight timeline due to the fast-paced technological environment. Hence, shortcuts as suggested by the authors of intervention mapping [15] had to be taken, similar to other projects applying this approach [106,107]. These shortcuts led to limited stakeholder involvement [107], especially in the planning group [15]; the focus on individual behavior change in women post-GDM without targeting the behavior of environmental agents such as their partners [108]; BMI-based tailoring only; and a respective prioritization of performance objectives, change objectives, behavior change methods, and practical applications. Lastly, we conducted intervention mapping steps 1 to 4 only and thus lack the planning and data on large-scale implementation (step 5), evaluation (step 6), and cost-effectiveness of the TRIANGLE intervention.

The user study sample was rather homogenous and may exhibit a socioeconomic bias due to its inclusion of iPhone users only and women with a high education level who were mostly native speakers. Women during the first 5 years post-GDM participated voluntarily, which may have preselected those with a higher motivation to test the TRIANGLE app. Therefore, the sample’s usage of the app, respective feedback, and suggested adaptations

may not be representative of the priority population. In addition, the limited content available for the user study may have biased the distribution of activities per lifestyle area. Future studies on the TRIANGLE app need to address these factors and test its usability on the Android system.

Options for Improvements of the TRIANGLE Intervention

In the scope of larger future projects, we suggest to additionally support behavior change by key environmental agents such as partners [108], health care professionals [94], or families [109] and to develop specific content for cultural subgroups that could not be addressed in this project [110]. Further, other cardiometabolic risk behaviors of women during the first 5 years post-GDM that we excluded need to be addressed, such as breastfeeding [111] and smoking [112]. In addition, new strategies associated with health benefits for women during the first 5 years post-GDM that contribute to the existing behavioral goals should be incorporated, such as intuitive eating [113] or postpartum emotional health [102]. Next, we identified other subgroups that need to receive tailored content in the future, such as those with low socioeconomic status [26]. Untapped behavior change methods may be added, such as motivational interviewing. Motivational interviewing was recently implemented in a digital lifestyle modification program [114] and is promising for future mHealth interventions. In general, the craving for more content needs to be addressed (eg, by weekly new content cycles [93]). Lastly, new features that further increase engagement with the app will have to be added and some of the coaching needs to be automated.

Conclusions

The intervention mapping steps 1 to 4 were used to plan, develop, and pretest the digital TRIANGLE intervention to prevent cardiometabolic disease in women during the first 5 years post-GDM. This project description with a detailed intervention model extends previous programs post-GDM and may guide the development of similar complex smartphone solutions to support behavior change. The TRIANGLE intervention uses one of the first interactive smartphone apps to address individual habit changes in several lifestyle areas with diverse behavior change methods and flexible pacing. The TRIANGLE app meets academic and industrial standards as a result of a respective cooperation. A user sample from the target group accepted the novel TRIANGLE intervention, rated the TRIANGLE iOS app to be of high quality, and perceived it to have a large impact on behavior. The TRIANGLE app's core features of “challenge system,” “human coaching,” and “library” seem suitable for large-scale implementation and the overall feedback was positive. However, some additional user needs and subgroups will have to be addressed to further refine the TRIANGLE intervention. The Test TRIANGLE study (German clinical trials register DRKS00012996), a small multicenter randomized controlled clinical trial, explores whether the TRIANGLE intervention is followed and accepted by women during the first years post-GDM and whether it changes their cardiometabolic risk behaviors. After the programming of the app for the Android system, a larger clinical trial needs to assess the program’s efficacy and cost-effectiveness in a representative sample. Intervention mapping steps 5 and 6 may guide program implementation and evaluation in a routine care setting, with increased stakeholder involvement during the respective planning.

Acknowledgments

This project, including the user study with the TRIANGLE app, was funded by the Else Kröner-Fresenius-Stiftung. The funders had no role in the development process, study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are grateful to all participants in the TRIANGLE user study and to the diabetes care team of the Medizinische Klinik und Poliklinik IV. We thank the following external partners for their contributions: Studio Bodyconcept, Linova Software GmbH, QuickBird Studios GmbH, Stefan Kofler (lead software developer), Dr Lukas Alperowitz (technical project manager), Elisabeth Wagner (logo design), Dr Sabine Waadt-Heim (psychological content), Dr Gabriele Duran-Atzinger (psychological content), Amelie Mueller (psychological content), and Kersten Hüttner (cameraman).

Conflicts of Interest

None declared.

Multimedia Appendix 1
TRIANGLE app content per lifestyle area as tested in the user study.
[DOCX File , 41 KB - mhealth_v9i5e26163_app1.docx]

Multimedia Appendix 2
Mixed methods design of the TRIANGLE user study.
[DOCX File , 38 KB - mhealth_v9i5e26163_app2.docx]

Multimedia Appendix 3
Test setup of the think-aloud sessions in the TRIANGLE user study.

[DOCX File , 86 KB - mhealth_v9i5e26163_app3.docx]

Multimedia Appendix 4
Specification of the TRIANGLE app user logs.
[DOCX File , 39 KB - mhealth_v9i5e26163_app4.docx]

Multimedia Appendix 5
Logic model of the problem of type 2 diabetes and related cardiometabolic disturbances after gestational diabetes.
[DOCX File , 132 KB - mhealth_v9i5e26163_app5.docx]

Multimedia Appendix 6
Performance objectives for behavioral outcomes.
[DOCX File , 42 KB - mhealth_v9i5e26163_app6.docx]

Multimedia Appendix 7
Logic model of behavior change for cardiometabolic health after gestational diabetes.
[DOCX File , 185 KB - mhealth_v9i5e26163_app7.docx]

Multimedia Appendix 8
TRIANGLE intervention modules and submodules.
[DOCX File , 40 KB - mhealth_v9i5e26163_app8.docx]

References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
2. Vounzoulaki E, Khunti K, Abner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. BMJ 2020 May 13;369:m1361-m1311 [FREE Full text] [doi: 10.1136/bmj.m1361] [Medline: 32404325]
3. Kaul P, Savu A, Nerenberg KA, Donovan LE, Chik CL, Ryan EA, et al. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: a population-level analysis. Diabet Med 2015 Mar;32(2):164-173. [doi: 10.1111/dme.12635] [Medline: 25407209]
4. Xu Y, Shen S, Sun L, Yang H, Jin B, Cao X. Metabolic syndrome risk after gestational diabetes: a systematic review and meta-analysis. PLoS One 2014;9(1):e87863 [FREE Full text] [doi: 10.1371/journal.pone.0087863] [Medline: 24498216]
5. Ratner RE, Christophi CA, Metzger BE, Dabelea D, Bennett PH, Pi-Sunyer X, Diabetes Prevention Program Research Group. Prevention of diabetes in women with a history of gestational diabetes: effects of metformin and lifestyle interventions. J Clin Endocrinol Metab 2008 Dec;93(12):4774-4779 [FREE Full text] [doi: 10.1210/jc.2008-0772] [Medline: 18826999]
6. Hewage SS, Wu S, Neelakantan N, Yoong J. Systematic review of effectiveness and cost-effectiveness of lifestyle interventions to improve clinical diabetes outcome measures in women with a history of GDM. Clin Nutr ESPEN 2020 Feb;35:20-29. [doi: 10.1016/j.clnesp.2019.10.011] [Medline: 31987117]
7. Goveia P, Cañon-Montañez W, Santos DDP, Lopes GW, Ma RCW, Duncan BB, et al. Lifestyle Intervention for the Prevention of Diabetes in Women With Previous Gestational Diabetes Mellitus: A Systematic Review and Meta-Analysis. Front Endocrinol (Lausanne) 2018;9:583 [FREE Full text] [doi: 10.3389/fendo.2018.00583] [Medline: 30344509]
8. Dennison RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on lifestyle changes to reduce the risk of developing Type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for practice. Diabet Med 2019 Jun;36(6):702-717 [FREE Full text] [doi: 10.1111/dme.13926] [Medline: 30723968]
9. Jones EJ, Fraley HE, Mazzawi J. Appreciating Recent Motherhood and Culture: A Systematic Review of Multimodal Postpartum Lifestyle Interventions to Reduce Diabetes Risk in Women with Prior Gestational Diabetes. Matern Child Health J 2017 Jan;21(1):45-57. [doi: 10.1007/s10995-016-2092-z] [Medline: 27435732]
10. Infanti JJ, O'Dea A, Gibson I, McGuire BE, Newell J, Glynn LG, et al. Reasons for participation and non-participation in a diabetes prevention trial among women with prior gestational diabetes mellitus (GDM). BMC Med Res Methodol 2014 Jan 24;14:13 [FREE Full text] [doi: 10.1186/1471-2288-14-13] [Medline: 24461045]
11. Zehle K, Smith BJ, Chey T, McLean M, Bauman AE, Cheung NW. Psychosocial factors related to diet among women with recent gestational diabetes: opportunities for intervention. Diabetes Educ 2008;34(5):807-814. [doi: 10.1177/0145721708323641] [Medline: 18832285]

12. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: A systematic review. Diabetes Res Clin Pract 2020 Apr;162:108084. [doi: 10.1016/j.diabres.2020.108084] [Medline: 32061819]
13. Mildon A, Sellen D. Use of mobile phones for behavior change communication to improve maternal, newborn and child health: a scoping review. J Glob Health 2019 Dec 1;9(2):020425 [FREE Full text] [doi: 10.7189/jogh.09.020425] [Medline: 31893032]
14. O'Reilly SL, Laws R. Health-e mums: Evaluating a smartphone app design for diabetes prevention in women with previous gestational diabetes. Nutr Diet 2019 Nov;76(5):507-514. [doi: 10.1111/1747-0080.12461] [Medline: 30109762]
15. Eldredge L, Markham C, Ruiter R, Kok G, Fernandez M, Parcel G. Planning health promotion programs: an intervention mapping approach. 4th ed. San Francisco: Jossey-Bass; 2016.
16. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40.
17. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
18. Virzi RA. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? Hum Factors 2016 Nov 23;34(4):457-468. [doi: 10.1177/001872089203400407]
19. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. New York, United States: Association for Computing Machinery; 1993 Presented at: The INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; 1993; Amsterdam, the Netherlands p. 206-213. [doi: 10.1145/169059.169166]
20. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Rottenkolber M, Ferrari U, Holland L, Aertsen S, Kammer NN, Hetterich H, et al. The Diabetes Risk Phenotype of Young Women With Recent Gestational Diabetes. J Clin Endocrinol Metab 2015 Jun;100(6):E910-E918. [doi: 10.1210/jc.2014-3898] [Medline: 25742512]
23. Gar C, Rottenkolber M, Haenelt M, Potzel AL, Kern-Matschilles S, Then C, et al. Altered metabolic and hormonal responses to moderate exercise in overweight/obesity. Metabolism 2020 Jun;107:154219. [doi: 10.1016/j.metabol.2020.154219] [Medline: 32240726]
24. Ferrari U, Künzel H, Tröndle K, Rottenkolber M, Kohn D, Fugmann M, et al. Poor sleep quality is associated with impaired glucose tolerance in women after gestational diabetes. J Psychiatr Res 2015 Jun;65:166-171. [doi: 10.1016/j.jpsychires.2015.02.012] [Medline: 25930074]
25. Ferrari U, Banning F, Freibothe I, Tröndle K, Sacco V, Wichmann C, et al. Depressive symptoms, impaired glucose metabolism, high visceral fat, and high systolic blood pressure in a subgroup of women with recent gestational diabetes. J Psychiatr Res 2018 Feb;97:89-93. [doi: 10.1016/j.jpsychires.2017.12.001] [Medline: 29223020]
26. Anna V, van der Ploeg HP, Cheung NW, Huxley RR, Bauman AE. Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. Diabetes Care 2008 Dec;31(12):2288-2293 [FREE Full text] [doi: 10.2337/dc08-1038] [Medline: 18809630]
27. Huy C, Loerbroks A, Hornemann A, Röhrig S, Schneider S. Prevalence, Trend and Determining Factors of Gestational Diabetes in Germany. Geburtshilfe Frauenheilkd 2012 Apr;72(4):311-315 [FREE Full text] [doi: 10.1055/s-0031-1298390] [Medline: 25284837]
28. Lim K, Chi C, Chan S, Lim SL, Ang SM, Yoong JS, et al. Smart Phone APP to Restore Optimal Weight (SPAROW): protocol for a randomised controlled trial for women with recent gestational diabetes. BMC Public Health 2019 Oct 15;19(1):1287 [FREE Full text] [doi: 10.1186/s12889-019-7691-3] [Medline: 31615456]
29. Lindström J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, Finnish Diabetes Prevention Study Group. The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care 2003 Dec;26(12):3230-3236. [doi: 10.2337/diacare.26.12.3230] [Medline: 14633807]
30. Retnakaran R. Hyperglycemia in pregnancy and its implications for a woman's future risk of cardiovascular disease. Diabetes Res Clin Pract 2018 Nov;145:193-199. [doi: 10.1016/j.diabres.2018.04.008] [Medline: 29679623]
31. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 2009 Nov 14;374(9702):1677-1686 [FREE Full text] [doi: 10.1016/S0140-6736(09)61457-4] [Medline: 19878986]
32. Moon JH, Kwak SH, Jung HS, Choi SH, Lim S, Cho YM, et al. Weight Gain and Progression to Type 2 Diabetes in Women With a History of Gestational Diabetes Mellitus. J Clin Endocrinol Metab 2015 Sep;100(9):3548-3555. [doi: 10.1210/JC.2015-1113] [Medline: 26171796]
33. Laaksonen MA, Knekt P, Rissanen H, Härkänen T, Virtala E, Marniemi J, et al. The relative importance of modifiable potential risk factors of type 2 diabetes: a meta-analysis of two cohorts. Eur J Epidemiol 2010 Mar;25(2):115-124. [doi: 10.1007/s10654-009-9405-0] [Medline: 20012885]

34. Emdin C, Anderson S, Woodward M, Rahimi K. Usual Blood Pressure and Risk of New-Onset Diabetes: Evidence From 4.1 Million Adults and a Meta-Analysis of Prospective Studies. J Am Coll Cardiol 2015 Oct 06;66(14):1552-1562 [FREE Full text] [doi: 10.1016/j.jacc.2015.07.059] [Medline: 26429079]
35. Jeppesen J, Hansen TW, Rasmussen S, Ibsen H, Torp-Pedersen C, Madsbad S. Insulin resistance, the metabolic syndrome, and risk of incident cardiovascular disease: a population-based study. J Am Coll Cardiol 2007 May 29;49(21):2112-2119 [FREE Full text] [doi: 10.1016/j.jacc.2007.01.088] [Medline: 17531661]
36. Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, et al. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. Diabetes Care 2016 May;39(5):833-846 [FREE Full text] [doi: 10.2337/dc15-2251] [Medline: 27208380]
37. Carr DB, Utzschneider KM, Hull RL, Tong J, Wallace TM, Kodama K, et al. Gestational diabetes mellitus increases the risk of cardiovascular disease in women with a family history of type 2 diabetes. Diabetes Care 2006 Sep;29(9):2078-2083. [doi: 10.2337/dc05-2482] [Medline: 16936156]
38. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 2003 Mar;26(3):725-731. [doi: 10.2337/diacare.26.3.725] [Medline: 12610029]
39. Halkoaho A, Kavilo M, Pietilä AM, Huopio H, Sintonen H, Heinonen S. Does gestational diabetes affect women's health-related quality of life after delivery? Eur J Obstet Gynecol Reprod Biol 2010 Jan;148(1):40-43. [doi: 10.1016/j.ejogrb.2009.09.025] [Medline: 19883969]
40. Azami M, Badfar G, Soleymani A, Rahmati S. The association between gestational diabetes and postpartum depression: A systematic review and meta-analysis. Diabetes Res Clin Pract 2019 Mar;149:147-155. [doi: 10.1016/j.diabres.2019.01.034] [Medline: 30735772]
41. Kolotkin RL, Andersen JR. A systematic review of reviews: exploring the relationship between obesity, weight loss and health-related quality of life. Clin Obes 2017 Oct;7(5):273-289 [FREE Full text] [doi: 10.1111/cob.12203] [Medline: 28695722]
42. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
43. Saboya PP, Bodanese LC, Zimmermann PR, Gustavo ADS, Assumpção CM, Londero F. Metabolic syndrome and quality of life: a systematic review. Rev Lat Am Enfermagem 2016 Nov 28;24:e2848 [FREE Full text] [doi: 10.1590/1518-8345.1573.2848] [Medline: 27901223]
44. Trikkalinou A, Papazafiropoulou AK, Melidonis A. Type 2 diabetes and quality of life. World J Diabetes 2017 Apr 15;8(4):120-129 [FREE Full text] [doi: 10.4239/wjd.v8.i4.120] [Medline: 28465788]
45. Grandy S, Chapman RH, Fox KM, SHIELD Study Group. Quality of life and depression of people living with type 2 diabetes mellitus and those at low and high risk for type 2 diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Int J Clin Pract 2008 Apr;62(4):562-568 [FREE Full text] [doi: 10.1111/j.1742-1241.2008.01703.x] [Medline: 18266708]
46. Rouhani MH, Haghighatdoost F, Surkan PJ, Azadbakht L. Associations between dietary energy density and obesity: A systematic review and meta-analysis of observational studies. Nutrition 2016 Oct;32(10):1037-1047. [doi: 10.1016/j.nut.2016.03.017] [Medline: 27238958]
47. Gabriel da Silva LB, Rosado EL, de Carvalho Padilha P, Dias JR, Moreira TM, de Paula TP, et al. Food intake of women with gestational diabetes mellitus, in accordance with two methods of dietary guidance: a randomised controlled clinical trial. Br J Nutr 2019 Jan;121(1):82-92. [doi: 10.1017/S0007114518001484] [Medline: 30392472]
48. Kim S, Kim M, Yang J, Park S, Yim CH, Han KO, et al. Nutritional risk factors of early development of postpartum prediabetes and diabetes in women with gestational diabetes mellitus. Nutrition 2011;27(7-8):782-788. [doi: 10.1016/j.nut.2010.08.019] [Medline: 21106349]
49. Davis JN, Shearrer GE, Tao W, Hurston SR, Gunderson EP. Dietary variables associated with substantial postpartum weight retention at 1-year among women with GDM pregnancy. BMC Obes 2017;4:31 [FREE Full text] [doi: 10.1186/s40608-017-0166-0] [Medline: 28794888]
50. Morrison MK, Koh D, Lowe JM, Miller YD, Marshall AL, Colyvas K, et al. Postpartum diet quality in Australian women following a gestational diabetes pregnancy. Eur J Clin Nutr 2012 Oct;66(10):1160-1165. [doi: 10.1038/ejcn.2012.84] [Medline: 22781022]
51. Sartorelli DS, Crivellenti LC, Zuccolotto DCC, Franco LJ. Relationship between minimally and ultra-processed food intake during pregnancy with obesity and gestational diabetes mellitus. Cad Saude Publica 2019 May 02;35(4):e00049318 [FREE Full text] [doi: 10.1590/0102-311X00049318] [Medline: 31066773]
52. D'Arcy E, Rayner J, Hodge A, Ross LJ, Schoenaker DAJM. The Role of Diet in the Prevention of Diabetes among Women with Prior Gestational Diabetes: A Systematic Review of Intervention and Observational Studies. J Acad Nutr Diet 2020 Jan;120(1):69-85.e7. [doi: 10.1016/j.jand.2019.07.021] [Medline: 31636052]
53. Marventano S, Salomone F, Godos J, Pluchinotta F, Del Rio D, Mistretta A, et al. Coffee and tea consumption in relation with non-alcoholic fatty liver and metabolic syndrome: A systematic review and meta-analysis of observational studies. Clin Nutr 2016 Dec;35(6):1269-1281. [doi: 10.1016/j.clnu.2016.03.012] [Medline: 27060021]

54. Muckelbauer R, Sarganas G, Grüneis A, Müller-Nordhorn J. Association between water consumption and body weight outcomes: a systematic review. Am J Clin Nutr 2013 Aug;98(2):282-299. [doi: 10.3945/ajcn.112.055061] [Medline: 23803882]
55. Tobias DK, Zhang C, Chavarro J, Olsen S, Bao W, Bjerregaard AA, et al. Healthful dietary patterns and long-term weight change among women with a history of gestational diabetes mellitus. Int J Obes (Lond) 2016 Nov;40(11):1748-1753 [FREE Full text] [doi: 10.1038/ijo.2016.156] [Medline: 27569683]
56. Bennett WL, Liu S, Yeh H, Nicholson WK, Gunderson EP, Lewis CE, et al. Changes in weight and health behaviors after pregnancies complicated by gestational diabetes mellitus: the CARDIA study. Obesity (Silver Spring) 2013 Jun;21(6):1269-1275 [FREE Full text] [doi: 10.1002/oby.20133] [Medline: 23666593]
57. Tobias DK, Hu FB, Chavarro J, Rosner B, Mozaffarian D, Zhang C. Healthful dietary patterns and type 2 diabetes mellitus risk among women with a history of gestational diabetes mellitus. Arch Intern Med 2012 Nov 12;172(20):1566-1572 [FREE Full text] [doi: 10.1001/archinternmed.2012.3747] [Medline: 22987062]
58. Gilbert L, Gross J, Lanzi S, Quansah DY, Puder J, Horsch A. How diet, physical activity and psychosocial well-being interact in women with gestational diabetes mellitus: an integrative review. BMC Pregnancy Childbirth 2019 Mar 07;19(1):60 [FREE Full text] [doi: 10.1186/s12884-019-2185-y] [Medline: 30732571]
59. Bao W, Tobias DK, Bowers K, Chavarro J, Vaag A, Grunnet LG, et al. Physical activity and sedentary behaviors associated with risk of progression from gestational diabetes mellitus to type 2 diabetes mellitus: a prospective cohort study. JAMA Intern Med 2014 Jul;174(7):1047-1055 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1795] [Medline: 24841449]
60. Persson M, Winkvist A, Mogren I. Lifestyle and health status in a sample of Swedish women four years after pregnancy: a comparison of women with a history of normal pregnancy and women with a history of gestational diabetes mellitus. BMC Pregnancy Childbirth 2015 Mar 13;15:57 [FREE Full text] [doi: 10.1186/s12884-015-0487-2] [Medline: 25884665]
61. Tindle HA, Duncan MS, Liu S, Kuller LH, Fugate Woods N, Rapp SR, et al. Optimism, pessimism, cynical hostility, and biomarkers of metabolic function in the Women's Health Initiative. J Diabetes 2018 Jun;10(6):512-523. [doi: 10.1111/1753-0407.12584] [Medline: 28703425]
62. Roest AM, Martens EJ, de Jonge P, Denollet J. Anxiety and risk of incident coronary heart disease: a meta-analysis. J Am Coll Cardiol 2010 Jun 29;56(1):38-46 [FREE Full text] [doi: 10.1016/j.jacc.2010.03.034] [Medline: 20620715]
63. Kuo W, Bratzke LC, Oakley LD, Kuo F, Wang H, Brown RL. The association between psychological stress and metabolic syndrome: A systematic review and meta-analysis. Obes Rev 2019 Nov;20(11):1651-1664. [doi: 10.1111/obr.12915] [Medline: 31347765]
64. Tsenkova VK, Karlamangla AS, Ryff CD. Parental History of Diabetes, Positive Affect, and Diabetes Risk in Adults: Findings from MIDUS. Ann Behav Med 2016 Dec;50(6):836-843 [FREE Full text] [doi: 10.1007/s12160-016-9810-z] [Medline: 27287937]
65. Chu P, Gotink RA, Yeh GY, Goldie SJ, Hunink MGM. The effectiveness of yoga in modifying risk factors for cardiovascular disease and metabolic syndrome: A systematic review and meta-analysis of randomized controlled trials. Eur J Prev Cardiol 2016 Mar;23(3):291-307. [doi: 10.1177/2047487314562741] [Medline: 25510863]
66. Pressman SD, Matthews KA, Cohen S, Martire LM, Scheier M, Baum A, et al. Association of enjoyable leisure activities with psychological and physical well-being. Psychosom Med 2009 Sep;71(7):725-732. [doi: 10.1097/PSY.0b013e3181ad7978] [Medline: 19592515]
67. Reutrakul S, Anothaisintawee T, Herring SJ, Balserak BI, Marc I, Thakkinstian A. Short sleep duration and hyperglycemia in pregnancy: Aggregate and individual patient data meta-analysis. Sleep Med Rev 2018 Aug;40:31-42. [doi: 10.1016/j.smrv.2017.09.003] [Medline: 29103944]
68. Anothaisintawee T, Reutrakul S, Van Cauter E, Thakkinstian A. Sleep disturbances compared to traditional risk factors for diabetes development: Systematic review and meta-analysis. Sleep Med Rev 2016 Dec;30:11-24. [doi: 10.1016/j.smrv.2015.10.002] [Medline: 26687279]
69. Aune D, Norat T, Romundstad P, Vatten LJ. Breastfeeding and the maternal risk of type 2 diabetes: a systematic review and dose-response meta-analysis of cohort studies. Nutr Metab Cardiovasc Dis 2014 Mar;24(2):107-115. [doi: 10.1016/j.numecd.2013.10.028] [Medline: 24439841]
70. Hoedjes M, Berks D, Vogel I, Franx A, Visser W, Duvekot JJ, et al. Effect of postpartum lifestyle interventions on weight loss, smoking cessation, and prevention of smoking relapse: a systematic review. Obstet Gynecol Surv 2010 Oct;65(10):631-652. [doi: 10.1097/OGX.0b013e3182077f64] [Medline: 21182803]
71. Wood W, Rünger D. Psychology of Habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
72. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: Modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
73. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
74. Petty R, Kasmer J, Haugtvedt C, Cacioppo J. Source and message factors in persuasion: A reply to stiff's critique of the elaboration likelihood model. Communication Monographs 1987 Sep;54(3):233-249. [doi: 10.1080/03637758709390229] [Medline: 1988]

75. Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
76. Prochaska JO, Redding CA, Harlow LL, Rossi JS, Velicer WF. The transtheoretical model of change and HIV prevention: a review. Health Educ Q 1994;21(4):471-486. [doi: 10.1177/109019819402100410] [Medline: 7843978]
77. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
78. Weinstein ND. The precaution adoption process. Health Psychol 1988;7(4):355-386. [doi: 10.1037//0278-6133.7.4.355] [Medline: 3049068]
79. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
80. Head KJ, Noar SM. Facilitating progress in health behaviour theory development and modification: the reasoned action approach as a case study. Health Psychol Rev 2014;8(1):34-52. [doi: 10.1080/17437199.2013.778165] [Medline: 25053006]
81. Parsons J, Sparrow K, Ismail K, Hunt K, Rogers H, Forbes A. A qualitative study exploring women's health behaviours after a pregnancy with gestational diabetes to inform the development of a diabetes prevention strategy. Diabet Med 2019 Feb;36(2):203-213. [doi: 10.1111/dme.13794] [Medline: 30098217]
82. Tang JW, Foster KE, Pumarino J, Ackermann RT, Peaceman AM, Cameron KA. Perspectives on prevention of type 2 diabetes after gestational diabetes: a qualitative study of Hispanic, African-American and White women. Matern Child Health J 2015 Jul;19(7):1526-1534. [doi: 10.1007/s10995-014-1657-y] [Medline: 25421329]
83. Zulfiqar T, Lithander F, Banwell C, Young R, Boisseau L, Ingle M, et al. Barriers to a healthy lifestyle post gestational-diabetes: An Australian qualitative study. Women Birth 2017 Aug;30(4):319-324. [doi: 10.1016/j.wombi.2016.12.003] [Medline: 28169159]
84. Kaiser B, Razurel C. Determinants of postpartum physical activity, dietary habits and weight loss after gestational diabetes mellitus. J Nurs Manag 2013 Jan;21(1):58-69. [doi: 10.1111/jonm.12006] [Medline: 23339495]
85. Kaiser B, Jeannot E, Razurel C. Determinants of Health Behaviors After Gestational Diabetes Mellitus: A Prospective Cohort Study in Geneva. J Midwifery Womens Health 2016 Sep;61(5):571-577. [doi: 10.1111/jmwh.12486] [Medline: 27434581]
86. Lie MLS, Hayes L, Lewis-Barned NJ, May C, White M, Bell R. Preventing type 2 diabetes after gestational diabetes: women's experiences and implications for diabetes prevention interventions. Diabet Med 2013 Aug;30(8):986-993. [doi: 10.1111/dme.12206] [Medline: 23534548]
87. Lim S, Dunbar JA, Versace VL, Janus E, Wildey C, Skinner T, et al. Comparing a telephone- and a group-delivered diabetes prevention program: Characteristics of engaged and non-engaged postpartum mothers with a history of gestational diabetes. Diabetes Res Clin Pract 2017 Apr;126:254-262. [doi: 10.1016/j.diabres.2017.02.026] [Medline: 28288435]
88. Razee H, van der Ploeg HP, Blignault I, Smith BJ, Bauman AE, McLean M, et al. Beliefs, barriers, social support, and environmental influences related to diabetes risk behaviours among women with a history of gestational diabetes. Health Promot J Aust 2010 Aug 01;21(2):130-137. [doi: 10.1071/HE10130] [Medline: 20701563]
89. O'Reilly SL, Dunbar JA, Versace V, Janus E, Best JD, Carter R, MAGDA Study Group. Mothers after Gestational Diabetes in Australia (MAGDA): A Randomised Controlled Trial of a Postnatal Diabetes Prevention Program. PLoS Med 2016 Jul;13(7):e1002092 [FREE Full text] [doi: 10.1371/journal.pmed.1002092] [Medline: 27459502]
90. de Bruin M, Sheeran P, Kok G, Hiemstra A, Prins JM, Hospers HJ, et al. Self-regulatory processes mediate the intention-behavior relation for adherence and exercise behaviors. Health Psychol 2012 Nov;31(6):695-703. [doi: 10.1037/a0027425] [Medline: 22390738]
91. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RAC, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
92. Sukumar N, Dallosso H, Saravanan P, Yates T, Telling C, Shorthose K, et al. Baby Steps - a structured group education programme with accompanying mobile web application designed to promote physical activity in women with a history of gestational diabetes: study protocol for a randomised controlled trial. Trials 2018 Dec 12;19(1):682 [FREE Full text] [doi: 10.1186/s13063-018-3067-8] [Medline: 30541621]
93. Nicklas JM, Leiferman JA, Lockhart S, Daly KM, Bull SS, Barbour LA. Development and Modification of a Mobile Health Program to Promote Postpartum Weight Loss in Women at Elevated Risk for Cardiometabolic Disease: Single-Arm Pilot Study. JMIR Form Res 2020 Apr 09;4(4):e16151 [FREE Full text] [doi: 10.2196/16151] [Medline: 32271149]
94. Nielsen KK, Dahl-Petersen IK, Jensen DM, Ovesen P, Damm P, Jensen NH, Face-it Study Group. Protocol for a randomised controlled trial of a co-produced, complex, health promotion intervention for women with prior gestational diabetes and their families: the Face-it study. Trials 2020 Mar 07;21(1):146 [FREE Full text] [doi: 10.1186/s13063-020-4062-4] [Medline: 32033613]
95. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention Mapping: Theory- and Evidence-Based Health Promotion Program Planning: Perspective and Examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]

96. Guo J, Chen J, Whittemore R, Whitaker E. Postpartum Lifestyle Interventions to Prevent Type 2 Diabetes Among Women with History of Gestational Diabetes: A Systematic Review of Randomized Clinical Trials. J Womens Health (Larchmt) 2016 Jan;25(1):38-49. [doi: 10.1089/jwh.2015.5262] [Medline: 26700931]
97. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Mar 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
98. Lipscombe LL, Delos-Reyes F, Glenn AJ, de Sequeira S, Liang X, Grant S, et al. The Avoiding Diabetes After Pregnancy Trial in Moms Program: Feasibility of a Diabetes Prevention Program for Women With Recent Gestational Diabetes Mellitus. Can J Diabetes 2019 Dec;43(8):613-620. [doi: 10.1016/j.jcjd.2019.08.019] [Medline: 31669188]
99. Brown J, Alwan NA, West J, Brown S, McKinlay CJ, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017 May 04;5:CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
100. O'Reilly S. Prevention of Diabetes after Gestational Diabetes: Better Translation of Nutrition and Lifestyle Messages Needed. Healthcare (Basel) 2014 Nov 21;2(4):468-491 [FREE Full text] [doi: 10.3390/healthcare2040468] [Medline: 27429288]
101. Nicklas JM, Zera CA, England LJ, Rosner BA, Horton E, Levkoff SE, et al. A web-based lifestyle intervention for women with recent gestational diabetes mellitus: a randomized controlled trial. Obstet Gynecol 2014 Sep;124(3):563-570 [FREE Full text] [doi: 10.1097/AOG.0000000000000420] [Medline: 25162257]
102. Athavale P, Thomas M, Delgadillo-Duenas AT, Leong K, Najmabadi A, Harleman E, et al. Linking High Risk Postpartum Women with a Technology Enabled Health Coaching Program to Reduce Diabetes Risk and Improve Wellbeing: Program Description, Case Studies, and Recommendations for Community Health Coaching Programs. J Diabetes Res 2016;2016:4353956 [FREE Full text] [doi: 10.1155/2016/4353956] [Medline: 27830157]
103. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Mar;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
104. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123.
105. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
106. Fassier JB, Sarnin P, Rouat S, Péron J, Kok G, Letrilliart L, et al. Interventions Developed with the Intervention Mapping Protocol in Work Disability Prevention: A Systematic Review of the Literature. J Occup Rehabil 2019 Mar;29(1):11-24. [doi: 10.1007/s10926-018-9776-8] [Medline: 29948471]
107. Majid U, Kim C, Cako A, Gagliardi AR. Engaging stakeholders in the co-development of programs or interventions using Intervention Mapping: A scoping review. PLoS One 2018;13(12):e0209826 [FREE Full text] [doi: 10.1371/journal.pone.0209826] [Medline: 30586425]
108. Kragelund Nielsen K, Groth Grunnet L, Terkildsen Maindal H, Danish Diabetes Academy WorkshopWorkshop Speakers. Prevention of Type 2 diabetes after gestational diabetes directed at the family context: a narrative review from the Danish Diabetes Academy symposium. Diabet Med 2018 Jun;35(6):714-720. [doi: 10.1111/dme.13622] [Medline: 29543341]
109. Pedersen ALW, Terkildsen Maindal H, Juul L. How to prevent type 2 diabetes in women with previous gestational diabetes? A systematic review of behavioural interventions. Prim Care Diabetes 2017 Oct;11(5):403-413. [doi: 10.1016/j.pcd.2017.05.002] [Medline: 28601549]
110. Mukerji G, Kainth S, Pendrith C, Lowe J, Feig DS, Banerjee AT, et al. Predictors of low diabetes risk perception in a multi-ethnic cohort of women with gestational diabetes mellitus. Diabet Med 2016 Oct;33(10):1437-1444. [doi: 10.1111/dme.13009] [Medline: 26499170]
111. Horta BL, Loret de Mola C, Victora CG. Long-term consequences of breastfeeding on cholesterol, obesity, systolic blood pressure and type 2 diabetes: a systematic review and meta-analysis. Acta Paediatr 2015 Dec;104(467):30-37. [doi: 10.1111/apa.13133] [Medline: 26192560]
112. Sattar N, Sorensen T, Taylor AE, Morris R, Munafò MR. Smoking and diabetes risk: building a causal case with clinical implications. Lancet Diabetes Endocrinol 2015 Dec;3(12):918-920. [doi: 10.1016/S2213-8587(15)00341-1] [Medline: 26388412]
113. Quansah DY, Gilbert L, Gross J, Horsch A, Puder JJ. Intuitive eating is associated with improved health indicators at 1-year postpartum in women with gestational diabetes mellitus. J Health Psychol 2019 Aug 22:1359105319869814. [doi: 10.1177/1359105319869814] [Medline: 31434517]
114. Coumans JMJ, Bolman CAW, Friederichs SAH, Oenema A, Lechner L. Development and Testing of a Personalized Web-Based Diet and Physical Activity Intervention Based on Motivational Interviewing and the Self-Determination Theory: Protocol for the MyLifestyleCoach Randomized Controlled Trial. JMIR Res Protoc 2020 Mar 04;9(2):e14491 [FREE Full text] [doi: 10.2196/14491] [Medline: 32014841]

Abbreviations

DPP: diabetes prevention program
GDM: gestational diabetes mellitus
MAGDA-DPP: Mothers After Gestational Diabetes in Australia Diabetes Prevention Program
mHealth: mobile health
PPSDiab: Prediction, Prevention and Sub-classification of type 2 Diabetes
SUS: System Usability Scale
uMARS: user version of the Mobile Application Rating Scale
V1: visit 1
V2: visit 2

Edited by L Buis; submitted 01.12.20; peer-reviewed by A Lake, W Na; comments to author 29.12.20; revised version received 11.01.21; accepted 19.03.21; published 11.05.21.

Please cite as:
Potzel AL, Gar C, Seissler J, Lechner A
A Smartphone App (TRIANGLE) to Change Cardiometabolic Risk Behaviors in Women Following Gestational Diabetes Mellitus: Intervention Mapping Approach
JMIR Mhealth Uhealth 2021;9(5):e26163
URL: https://mhealth.jmir.org/2021/5/e26163
doi:10.2196/26163
PMID:33973864

Original Paper

User-Centered Development of a Mobile App for Biopsychosocial Pain Assessment in Adults: Usability, Reliability, and Validity Study

Filipa Lopes[1], MSc; Mário Rodrigues[2], DPhil; Anabela G Silva[3], DPhil

[1]School of Health Sciences, University of Aveiro, Aveiro, Portugal

[2]Higher School of Technology and Management of Águeda, Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, Aveiro, Portugal

[3]Center for Health Technology and Services Research (CINTESIS.UA), School of Health Sciences, University of Aveiro, Aveiro, Portugal

Corresponding Author:
Anabela G Silva, DPhil
Center for Health Technology and Services Research (CINTESIS.UA), School of Health Sciences, University of Aveiro
Campus Universitário de Santiago
Aveiro, 3810-193
Portugal
Phone: 351 234247119 ext 27120
Email: asilva@ua.pt

Abstract

Background: Pain-related mobile apps targeting pain assessment commonly limit pain assessment to pain behaviors and physiological aspects. However, current guidelines state that pain assessment should follow the biopsychosocial model, clearly addressing biological, psychological, and social aspects of the pain experience. Existing reviews also highlight that pain specialists and end users are not commonly involved in the development process of mobile apps for pain assessment, negatively affecting the quality of the available apps.

Objective: This study aimed to develop a mobile app for pain assessment (AvaliaDor) and assess its usability, validity, reliability, and measurement error in a sample of real patients with chronic pain recruited from a physiotherapy clinic.

Methods: This study was divided into 2 phases: phase 1—development of the AvaliaDor app; and phase 2—assessment of the apps' usability, reliability, measurement error, and validity. AvaliaDor was developed (phase 1) based on the literature and the recommendations of physiotherapists and patients with pain in cycles of evaluation, inclusion of recommendations, and reevaluation until no further changes were required. The final version of the app was then tested in patients with musculoskeletal pain attending a private physiotherapy practice (phase 2) who were asked to use the app twice on 2 consecutive days for reliability purposes. In addition, participants had to complete a set of paper-based scales (Brief Pain Inventory, painDETECT, Pain Catastrophizing Scale, and Tampa Scale for Kinesiophobia), which were used to assess the validity (criterion validity and hypothesis testing) of the app, and the Post-Study System Usability Questionnaire was used to assess its usability.

Results: The development process (phase 1) included 5 physiotherapists external to the research team and 5 patients with musculoskeletal pain, and it resulted in the creation of an app named AvaliaDor, which includes an assessment of pain intensity, location, and phenotype; associated disability; and the issues of pain catastrophizing and fear of movement. A total of 52 patients with pain (mean age 50.12 years, SD 11.71 years; 39 females) participated in phase 2 and used the app twice. The Pearson correlation coefficient between the scores on the paper-based scales and the app ranged between 0.81 and 0.93 for criterion validity and between 0.41 and 0.59 for hypothesis testing. Test-retest reliability was moderate to good (intraclass correlation coefficient between 0.67 and 0.90) and the score for usability was 1.16 (SD 0.27), indicating good usability.

Conclusions: A mobile app named AvaliaDor was developed to assess the intensity, location, and phenotype of pain; associated disability; and the issues of pain catastrophizing and fear of movement in a user-centered design process. The app was shown to be usable, valid, and reliable for assessing pain from a biopsychosocial perspective in a heterogeneous group of patients with pain. Future work can explore the long-term use of AvaliaDor in clinical contexts and its advantages for the assessment and management of patients with pain.

(JMIR Mhealth Uhealth 2021;9(5):e25316) doi:10.2196/25316

KEYWORDS

pain assessment; mobile app; validity; reliability; usability; mHealth; pain; user-centered design

Introduction

Currently, there is a rocketing expansion in the development of mobile apps all over the world, largely due to the increased accessibility and global availability of smartphones [1]. The technology of mobile devices has greatly improved in recent years, with higher screen resolution and better processor performance, among other improvements in hardware. Furthermore, using smartphones to access the internet is part of everyday life [2], establishing a technological revolution. This advancement in digital technology is also changing health care [3] and making it more accessible [1].

The use of information and communication technologies in health is called eHealth [4]. According to the World Health Organization, mobile health (mHealth)—which covers medical and public health practices supported by mobile devices, such as mobile phones, user monitoring devices, personal digital assistants, and other wireless devices [5]—is a component of eHealth. Among the advantages of mHealth is its use of smartphones instead of traditional personal computers, which provides self-monitoring support to the user in most everyday situations [6,7]. mHealth facilitates access to information related to the user's health conditions or treatments, allows the organization and recording of health information, allows for self-monitoring and self-management, and facilitates communication between patients and health care providers and interaction between health care users or providers and health services registration systems [8]. Also, eHealth has been shown to overcome current health care problems and limitations, such as difficulties accessing timely and continuous health care, the associated costs, mobility limitations, or long wait times [9,10]. This is particularly relevant for chronic conditions, such as chronic pain, in which self-management is a key component of the intervention.

Pain-related mobile apps targeting pain assessment commonly include electronic diaries to monitor pain characteristics such as intensity, location, factors that aggravate or alleviate pain, and/or intake of medication [8,11,12], largely limiting pain assessment to pain behaviors and physiological aspects and not including a biopsychosocial assessment of pain [8]. A systematic search found 142 pain-related apps, of which 28 were primarily intended for pain assessment [8]. However, when comparing the existing mobile apps against the existing data on their development, it was also found that data on reliability, validity, and usability were scarce for pain assessment apps [8]. Another study corroborated these findings by concluding that despite the large and growing number of existing mobile apps related to pain, their quality is seldom guaranteed [13]. This may be due, in part, to the lack of involvement of health professionals and users in the process of developing the app [14]. This involvement is a way of ensuring the validity of the content of the mobile app [15] and that it meets the users' needs and requirements [16], increasing its likelihood of being usable. Usability is an essential criterion in the evaluation of mHealth apps [17,18] and has been defined as the extent to which a product can be used by users to achieve specific goals with effectiveness, efficiency, and satisfaction and in a specific context of use [19]. Nevertheless, a systematic review that synthesized and evaluated existing studies on the assessment of the usability of pain-related apps found 31 manuscripts on the usability of 32 pain apps [20]. The results of this systematic review suggested that several important methodological aspects regarding the assessment of usability are not being considered when developing pain-related apps, such as not using reliable and valid instruments to assess usability [20].

As measurement instruments, mobile apps that intend to be used for pain assessment need to provide evidence of their reliability, measurement error, and validity. Reliability refers to the consistency of results in repeated measurements performed in similar circumstances [21]. Measurement error informs on the amount of change needed between successive measurements that could be considered a real change in the person's condition [16]. Validity refers to whether an instrument measures what it intends to measure [21]. The main aim of this study was to develop a mobile app for the biopsychosocial assessment of pain and to assess it for usability, validity, reliability, and measurement error in a sample of real patients with chronic pain recruited from a physiotherapy clinic. The app was developed in a user-centered design paradigm, involving the users from the very beginning.

Methods

This study describes the development of a mobile app for pain assessment and its usability, validity, reliability, and measurement error assessment. It received ethical approval from the Council of Ethics and Deontology of the University of Aveiro. All participants provided written informed consent before entering the study.

Procedures

This study was divided into 2 phases: phase 1—development of the app; and phase 2—assessment of the app in terms of its usability, reliability, measurement error, and validity. The development of the app (phase 1) included the (1) validity of the app's content, (2) content analysis and user interface design, (3) prototype development, (4) prototype testing, and (5) prototype modification and construction of the final app. These phases and subphases are described below.

Phase 1: Development of the App

Validity of the App's Content

The first step in the development of the app was to define the fundamental aspects that should be considered for a biopsychosocial pain assessment. This was based on the literature on pain assessment, existing mobile apps for pain assessment, and input from physiotherapists and patients with chronic pain. First, we considered the recommendations from the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) on the pain-related domains that

should be assessed for patients with chronic pain [22] and analyzed existing apps, such as Keele Pain Recorder [23] and Pain Monitor [24]. Based on this analysis, and previous experience and knowledge, the 2 physiotherapists on the research team (FL and AGS) defined the preliminary contents that the app should cover: pain intensity, pain location, impact of pain, influence of pain on sleep, pain catastrophizing, fear of movement, and main mechanism of pain (eg, nociceptive or neuropathic). Then, validated instruments that covered most of the defined domains were identified and adapted to be included in the app: the Brief Pain Inventory-Short Form [25,26], covering pain intensity, disability, and sleep; and the painDETECT questionnaire [27], covering the phenotype of pain. Both scales have a body map for pain location. Also, catastrophizing and fear of movement were assessed in the mobile app with a single question on each of the constructs, retrieved from the Pain Catastrophizing Scale and the Tampa Scale for Kinesiophobia, respectively. The questions were identified based on their predictive ability [28,29].

Content Analysis and User Interface Design

Once the information to be included in the app was determined, paper-based mock-ups were produced using Adobe XD (Adobe Inc) to visualize the information displayed in a format similar to that of the proposed mobile app. These mock-ups were shown to 5 physiotherapists—experts in musculoskeletal physiotherapy—and 5 patients with musculoskeletal pain. Physiotherapists had to have a master's degree or be enrolled in their second year of a master's degree program in the area of musculoskeletal physiotherapy and have a minimum of 2 years of professional experience in musculoskeletal physiotherapy. Patients were included if they were aged ≥18 years, reported musculoskeletal pain, were able to speak and read Portuguese, and understood the main objectives of the study. Both physiotherapists and patients were asked the following questions: (1) Do you think the app covers all relevant content for pain assessment?; (2) Are the questions clear and concise?; (3) Are the response options adequate?; (4) Is the order of the questions the most appropriate?; and (5) Do you have any suggestions for improvement? All interviews were recorded, and notes were taken regarding suggested changes. We found that both patients and physiotherapists provided similar recommendations, and no new information was added after the fifth participant of each group (we reached theoretical saturation); thus, no further participants were recruited at this stage. The sample size for the development phase of the app was informed by data saturation. This is a principle widely accepted in qualitative research and indicates that, based on the data that have been collected, further data collection is unnecessary [30].

Prototype Development

The prototype was developed for the Android operating system using Android Developer Studio (Google, Inc). First, we implemented the screen layouts according to the mock-ups and tested them with distinct screen sizes, ratios, and pixel densities for mobile phones. Adjustments were made for better fits, and all adjustments were validated by the mock-up's designer. After layout validation, the elements were made functional and the dynamic look and feel were also validated by the mock-up's designer.

At the end of each registration, the data were sent via the internet to a server featuring a back office for data visualization and download. Each record was a tuple containing all collected answers plus images of the marked body maps in PNG format. If a network connection was not available, the data were stored locally—the tuple in an SQLite database and the images in the app-specific folder. Whenever locally stored records existed, the user was informed on the first screen of the app and he/she could choose to send them. The records sent were then erased from the local database and file system.

Prototype Testing: Preliminary Usability Evaluation

The information gathered in previous steps was used to inform the first version of the app—AvaliaDor, version 1. This version of the app was tested by the same 5 physiotherapists and 5 patients that had analyzed the mock-ups. They were interviewed once more after having used the app in a preliminary usability evaluation. The physiotherapists were given the app 1 week before the interview. Both physiotherapists and patients were asked the following questions: (1) Do you think the app is functional?; (2) Do you think it is simple to use?; (3) Do you notice any aspect that should be improved ?; (4) On what model of mobile phone did you use the app?; and (5) Was the app well-adjusted to the screen? In addition, the time taken to complete a full pain assessment using the app and any errors were registered (performance indicators for usability evaluation).

Prototype Modification and Construction of the Final App

The physiotherapists in the team and the programmer discussed the suggestions made in the previous step of app development, and another version of the app—AvaliaDor, version 2—was developed and tested again by the same 5 patients. Physiotherapists were not included at this stage because they made only minor recommendations for improvement in the prototype testing subphase. Once no further changes were recommended by patients, the AvaliaDor, version 2, became the final version of the app.

Phase 2: Assessment of the App in Terms of its Usability, Reliability, Measurement Error, and Validity

Participants and Instruments

Participants were patients with musculoskeletal pain attending a private physiotherapy practice. To enter the study, they were required to report pain in the limbs or back, be aged ≥18 years, be able to read and speak Portuguese, and understand the aims of the study (assessed by asking participants to explain in their own words what the study was about). The sample size was based on indicators for reliability studies. To attain a maximum intraclass correlation coefficient (ICC) of 0.8—with 2 measurements, an alpha of 5%, and a power at 80%—at least 46 participants were needed [31].

Participants were asked to use the AvaliaDor app to assess their pain on 2 consecutive days for reliability assessment purposes. On the second day, before using the app, patients were asked if there were changes in their pain since the previous day. Those

answering yes to this question were excluded from the reliability analysis. Also, in the first session, patients were asked to complete the following questionnaires, which were used to assess hypothesis testing and criterion validity:

- User characterization questionnaire—this questionnaire collected data on gender, age, educational qualifications, marital status, current clinical diagnoses, and duration of pain.
- Pain Catastrophizing Scale—consists of 13 items that depict thoughts, perceptions, or feelings associated with pain [32,33]. Participants are asked to indicate to what extent they present with the thoughts and feelings described using a 4-point Likert scale (0=never to 4=always). The final score represents the sum of all items and varies between 0 and 52, with higher scores reflecting higher levels of catastrophic thoughts [34,35]. The ICC for the reliability of the Portuguese version of this scale varies between 0.78 and 0.82 [32].
- Tampa Scale for Kinesiophobia—consists of 13 items that measure kinesiophobia under generic conditions. Each item is scored on a 4-point Likert scale, from 1 (strongly disagree) to 4 (strongly agree). The total score corresponds to the sum of the scores obtained for each of the items and varies from 13 to 52 points, where 13 represents the lowest and 52 the highest degree of kinesiophobia [36]. It showed good levels of test-retest reliability (ICCs between 0.94 and 0.98) and internal consistency with a Cronbach α of 0.82 [36].
- Brief Pain Inventory—consists of 2 subscales: the pain severity and pain interference subscales [25,32]. The score for pain severity ranges from 0 to 40 and the score for pain interference ranges from 0 to 70 [37]. This instrument has excellent internal consistency levels with Cronbach α values for the pain severity and interference subscales of 0.99 and 0.84, respectively. Its test-retest reliability also has very satisfactory values: ICCs of 0.88 for pain severity and 0.84 for pain interference [32].
- painDETECT questionnaire—aims to identify the nociceptive, neuropathic, or mixed component of the pain phenotype [27]. Its score ranges from 0 to 38, where 0 to 12 corresponds to nociceptive pain, 13 to 18 indicates a mixed pain, and 19 to 38 corresponds to neuropathic pain. This questionnaire has been shown to be reliable [38].

Also, further assessment of the app's usability was conducted through the Post-Study System Usability Questionnaire. This questionnaire was developed to assess users' satisfaction with the usability of a digital solution. It consists of 19 items that can either be scored using an 8-point Likert scale, anchored at the lower end with "totally agree" and at the higher end with "totally disagree," or be scored as not applicable. The total score is the mean of all items and lower scores are indicative of better usability. The Portuguese version has good internal consistency with a Cronbach α of 0.80 and acceptable interrater reliability (ICC 0.67) [39].

Statistical Analysis

SPSS software (version 26; IBM Corp) was used to perform data analysis. For sample characterization, the mean and standard deviation were used for continuous variables, while frequency distribution and percentage were used for ordinal and nominal variables. The criterion validity of the app questions on pain location, pain intensity, pain interference, and pain phenotype was assessed by correlating the app scores with the paper-based scores of the respective paper-based questionnaire using the Pearson correlation coefficient. Criterion validity was considered to exist when the correlation between the results of the app and the respective gold standard (paper-based questionnaire) was ≥0.7 [40]. For the questions on fear of movement and catastrophizing, construct validity was assessed by correlating the scores obtained using the app with the scores obtained on the paper-based Tampa Scale for Kinesiophobia and Pain Catastrophizing Scale, respectively, also using the Pearson correlation coefficient. For construct validity, the correlation coefficient was interpreted as indicating correlation—low (<0.3), moderate (0.3-0.5), and strong (>0.5) [41]—and we hypothesized that there would be a moderate correlation between the app scores and the paper-based scores.

Test-retest reliability was assessed using the ICC (bidirectional randomness, absolute agreement) and the respective 95% confidence interval. The results were interpreted as weak (ICC <0.50), moderate (ICC 0.50-0.75), good (ICC 0.75-0.90), and excellent (ICC ≥0.90) [42]. In addition, measurement error was assessed using the standard error of measurement (SEM) and the minimal detectable difference (MDD) with a confidence level of 95% (MDD95), calculated as $SEM = SD\sqrt{(1-ICC)}$ and $MDD95 = SEM \times 1.96 \times \sqrt{2}$ [43]. We also calculated the limits of agreement and constructed Bland-Altman plots [44] both for the criterion validity and for the test-retest reliability.

Results

Phase 1: Development of the AvaliaDor App

The most relevant characteristics of the patients with pain and the physiotherapists who contributed to the process of development and analysis of the app are presented in Tables 1 and 2, respectively.

Table 1. Characteristics of patients involved in the development phase (phase 1) of the mobile app (n=5).

Variable	Number of patients
Gender	
Female	4
Male	1
Age (years)	
18-30	1
31-50	2
51-60	1
61-70	1
Education level (school years)	
6	1
9	1
12	1
University	2
Location of main pain complaint	
Neck	1
Knee	1
Hip	1
Face	1
Low back	1

Table 2. Characteristics of the physiotherapists involved in the development phase (phase 1) of the mobile app (n=5).

Variable	Number of physiotherapists
Gender	
Female	3
Male	2
Age (years)	
25-30	2
31-40	3
Experience as a physiotherapist (years)	
5-9	3
10-12	2

The main suggestions of the physiotherapists regarding the mock-ups included creating a final summary report of the assessment, changing the caption of the body map, and adding an option regarding the location of the pain instead of featuring a body map. Only 1 patient suggested changes to the app, namely to change the order of the questions. Table 3 lists the patients' and physiotherapists' suggestions and comments, as well as the resulting decisions made by the research team.

Table 3. Summary of physiotherapists' and patients' comments on the mock-ups of the AvaliaDor app.

Question	Physiotherapists' comments	Patients' comments	Decisions made by the research team
Do you think the app covers all the relevant content for pain assessment?	All physiotherapists found that the app covered all relevant content, apart from 1 physiotherapist who suggested including injury mechanism, pain duration, relief, and/or worsening factors.	All patients stated that everything relevant was included.	As most physiotherapists and all users reported that the app covered the most relevant content, no further contents were included.
Are the questions clear and concise?	All physiotherapists stated that the questions were clear and concise. One physiotherapist noted that in the question, "Does your pain spread to other regions of the body?", an affirmative answer required the user to go back, and this should be amended.	All patients reported that the questions were clear and concise.	A response option was added following the question, "Does your pain spread to other regions of the body?"
Are the response options adequate?	Two physiotherapists, despite claiming that the response options were adequate, suggested that it should be possible for users to mark the exact area of pain on the body map and that response options should be wider apart so that the different options are clearer.	All users reported that the answers were adequate. One user even mentioned that the examples were very enlightening and illuminating.	Both of the physiotherapists' suggestions were included in the app.
Is the order of questions the most appropriate?	All physiotherapists agreed that the order was the most appropriate.	All but 1 patient agreed on the order of the questions. One patient suggested that the questions on pain interference should be listed before the questions on pain intensity and pain location.	As all physiotherapists and most patients agreed with the order of the questions, which is also aligned with what is more common in existing apps and scales, no change was made.
Do you have any suggestions for improvement?	A number of suggestions were made: include a final report with a summary score; calculate the average pain from the worst and least pain intensities; precede the question on the effect of the treatment with a question on whether patients were receiving treatment; include "injury mechanisms," pain duration, and "relieving/aggravating factors"; change the body map caption to include both chief pain complaint and other pain complaints; and check the verb tenses so that all questions use the same tense.	No suggestions for improvement were made.	A summary report was included; the question on the average pain intensity was maintained in line with the original version of the questionnaire; and, for a similar reason, a question on whether patients were receiving treatment was not included. The team decided not to include questions on the mechanism of injury, duration of pain, and relieving/aggravating factors because it would make the app questionnaire too extensive. The legend of the body map was modified in line with physiotherapists' comments, and the verb tenses were corrected.

The first version of the app was then developed and again shown to users for assessment. Physiotherapists' and patients' general opinion of the app was quite favorable, reporting that it was functional and easy to use. Criticisms and suggestions from physiotherapists and users are presented in Table 4. Table 4 also shows the models of the mobile phones used to run the app because the characteristics of the screen have an impact on the readability and activation of the response options, and the type of processor influences the fluidity of the rendering of the body map. Participants took between 5 and 8 minutes to complete all of the steps for pain assessment in the app, and no app-related errors were identified.

Next, a second version of the app was developed, and patients reported that no further changes were required.

The final version of the AvaliaDor app includes the following:

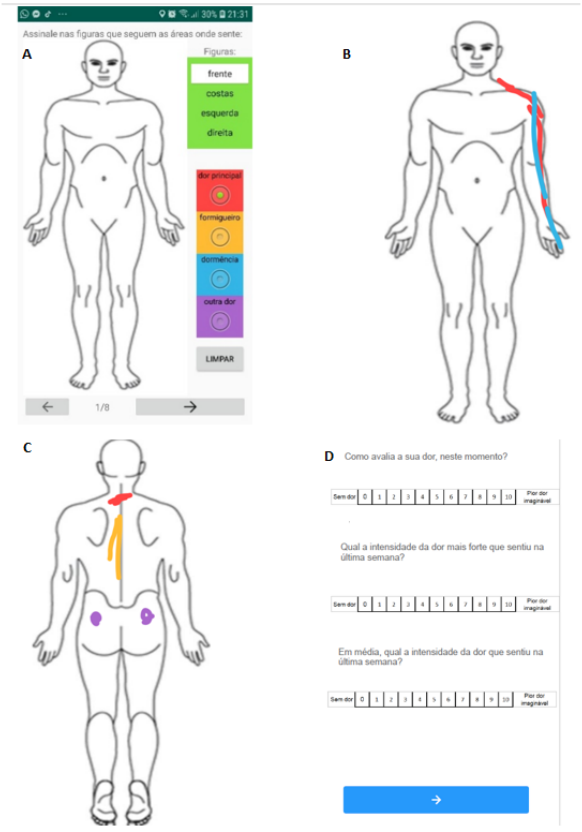
- a login page;
- a body map (front, back, right side, and left side) for the patient to draw pain, pins and needles, numbness, or another symptoms;
- questions on pain phenotype based on the painDETECT questionnaire;
- questions on pain intensity (at present, at its worst, and over the last week);
- questions on pain interference based on the Brief Pain Inventory;
- 1 question on fear of movement;
- 1 question on pain catastrophizing; and
- a report generated with the patient's data in the form of a Microsoft Excel file.

Table 4. Summary of physiotherapists' and patients' comments on the first version of the AvaliaDor app.

Commenter (models of mobile phone) and suggestions	Decisions made by the research team
Physiotherapists (Samsung Galaxy S7; Samsung Galaxy S6; Huawei P10 Lite; Huawei P20)	
The app was considered easy to use except for the body map, which might be too small for patients to mark different symptoms on (n=2).	The size of the body map was increased.
It was suggested that a question asking whether the patient felt some relief from their symptoms should be included before asking the patient to indicate the extent of that relief (n=2).	No question was added, as it was believed that the patient could indicate 0% improvement if no relief from symptoms was experienced.
The body map's legend should state "main pain" instead of "pain" to minimize confusion if the patient had more than one painful body site (n=2).	The body map's legend was changed accordingly.
There should be a "go back" sign (n=1).	An arrow was added to make it clearer how to go back.
The font size might be small for older adults (n=1).	No change to the font size was made, as it was not identified by the older patients in the sample as being a limitation.
Patients (Samsung Galaxy S7)	
It was difficult to draw the pain area (n=2).	The size of the body map was increased.
The researcher was asked how someone should answer the question about pain relief in the case that no treatment was being received (n=2).	No question was added, as it was considered that the patient could indicate 0% improvement if no pain relief was experienced.

Figure 1 shows the following screen excerpts from the app: the body map (Figure 1A); filled body maps (Figure 1B and 1C); and a page of questions (Figure 1D).

Figure 1. Print screens from the app. (A) Front body map. Front (B) and back (C) body maps after being filled in by the patient. (D) List of questions.



Phase 2: Usability, Reliability, Measurement Error, and Validity

The study sample consisted of 52 participants aged between 22 and 72 years (mean 50.12 years, SD 11.71 years). Table 5 presents a more detailed description of the sample.

Table 5. Characteristics of patients who participated in the assessment of the final version of the app (n=52).

Characteristics	Values, n (%)
Gender	
Female	39 (75)
Male	13 (25)
Age (years)	
<50	23 (44)
≥50	29 (56)
Education level (school years)	
4	5 (10)
6	19 (37)
9	12 (23)
12	4 (8)
University	12 (23)
Location of main pain complaint	
Neck	6 (12)
Low back	8 (15)
Shoulders	17 (33)
Elbows	4 (8)
Hands	3 (6)
Hips	1 (2)
Knees	4 (8)
Legs	1 (2)
Feet	8 (15)
Pain duration	
<3 months	7 (13)
≥3 to <6 months	8 (15)
≥6 months to <1 year	2 (4)
≥1 to <2 years	12 (23)
≥2 to <5 years	8 (15)
≥5 years	15 (29)

Usability Evaluation

Usability was tested using the Post-Study System Usability Questionnaire, and the average score obtained was 1.16 (SD 0.27), indicating that the app was considered usable.

Reliability and Measurement Error

The ICCs obtained between the 2 assessments and the mobile app indicates moderate to excellent reliability, with ICCs between 0.67 and 0.90, as presented in Table 6. Table 6 also presents the SEM and MDD. Visual inspection of the limits of Bland-Altman plots (Figures 2-7) shows a symmetrical distribution around the mean, close to 0, with no systematic or proportional bias.

Figure 2. Bland-Altman limits for the Brief Pain Inventory severity subscale between the 2 assessments using the app.

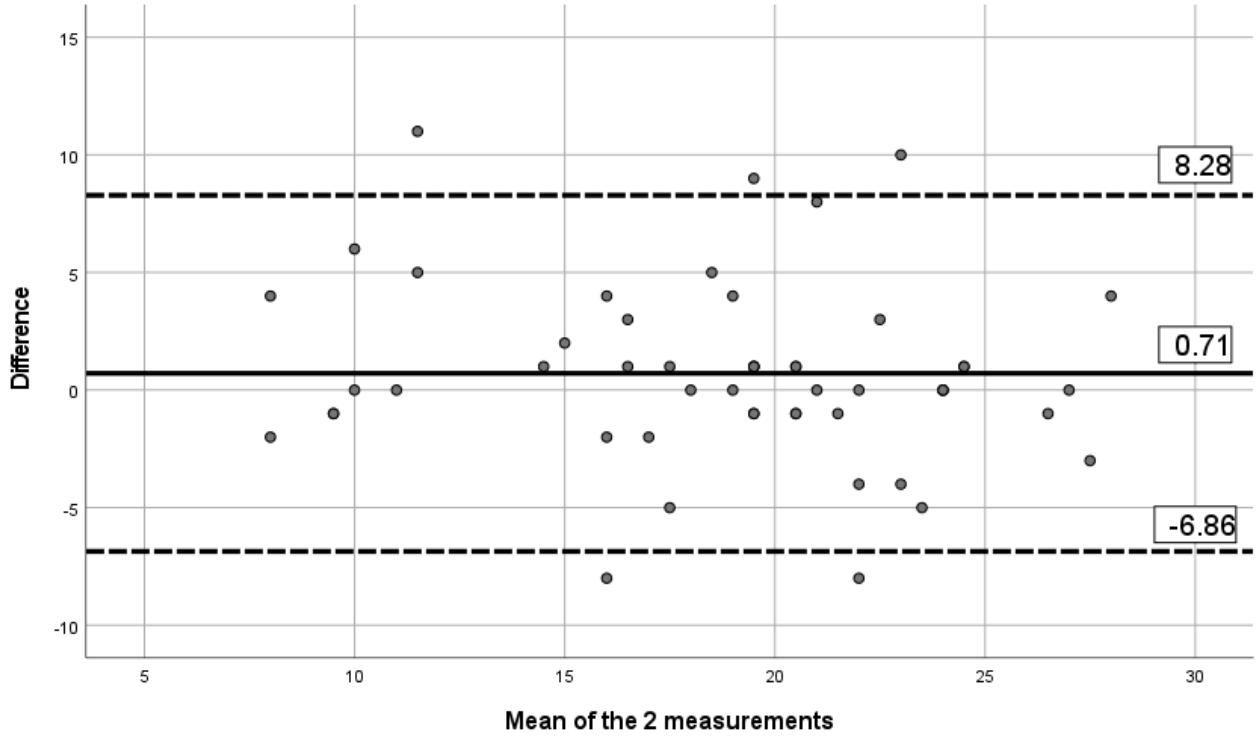


Figure 3. Bland-Altman limits for the Brief Pain Inventory pain interference subscale between the 2 assessments using the app.

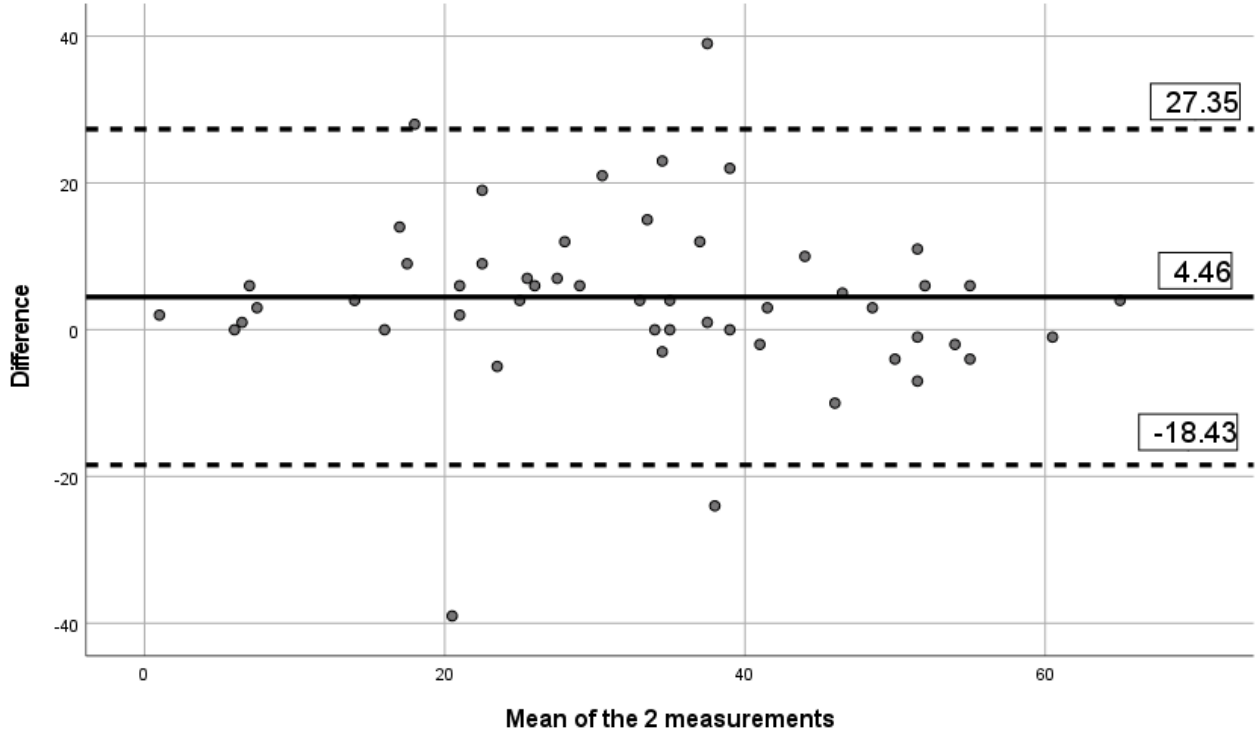


Figure 4. Bland-Altman limits for the painDETECT questionnaire between the 2 assessments using the app.

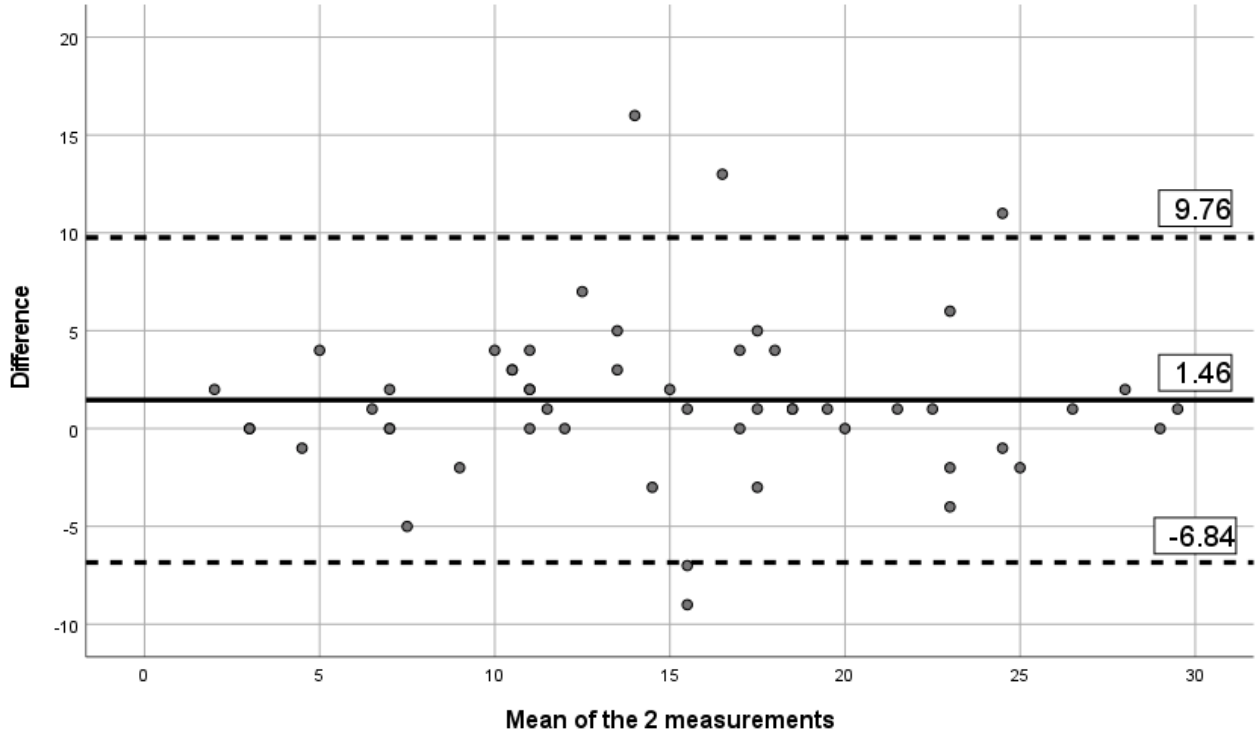


Figure 5. Bland-Altman limits for the issue of pain catastrophizing between the 2 assessments using the app.

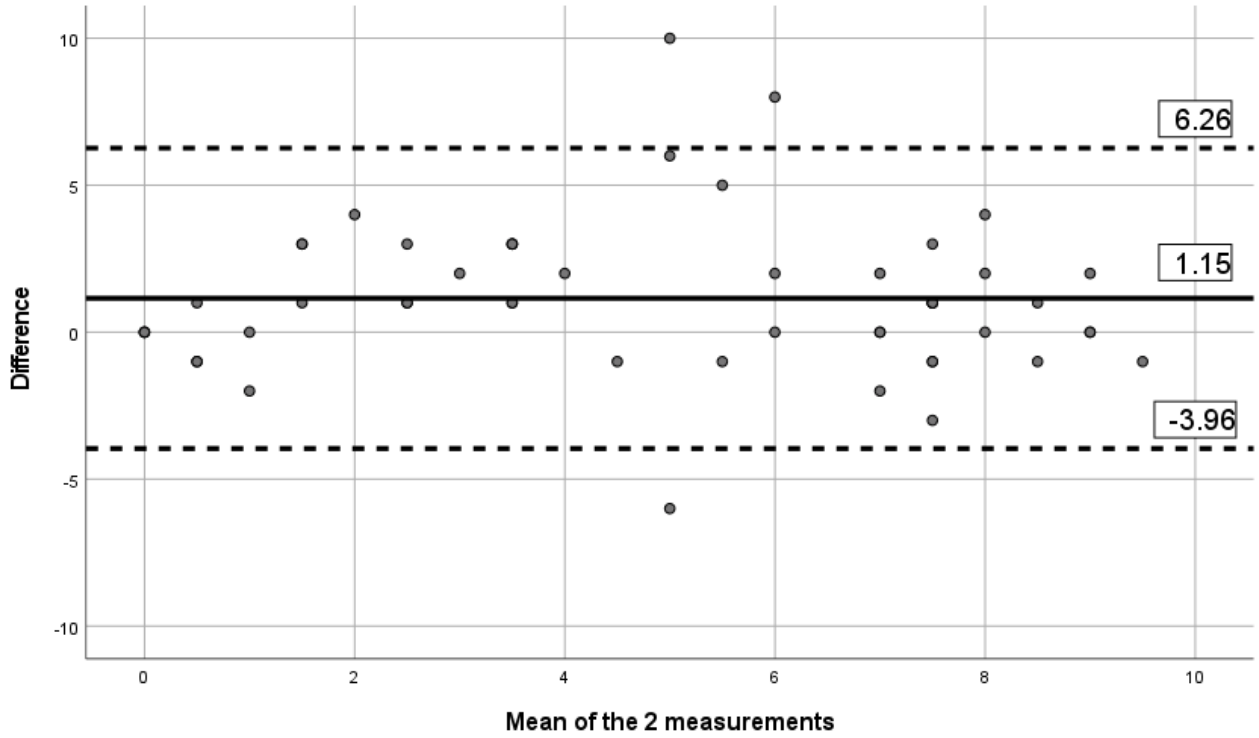


Figure 6. Bland-Altman limits for the issue of fear of movement between the 2 assessments using the app.

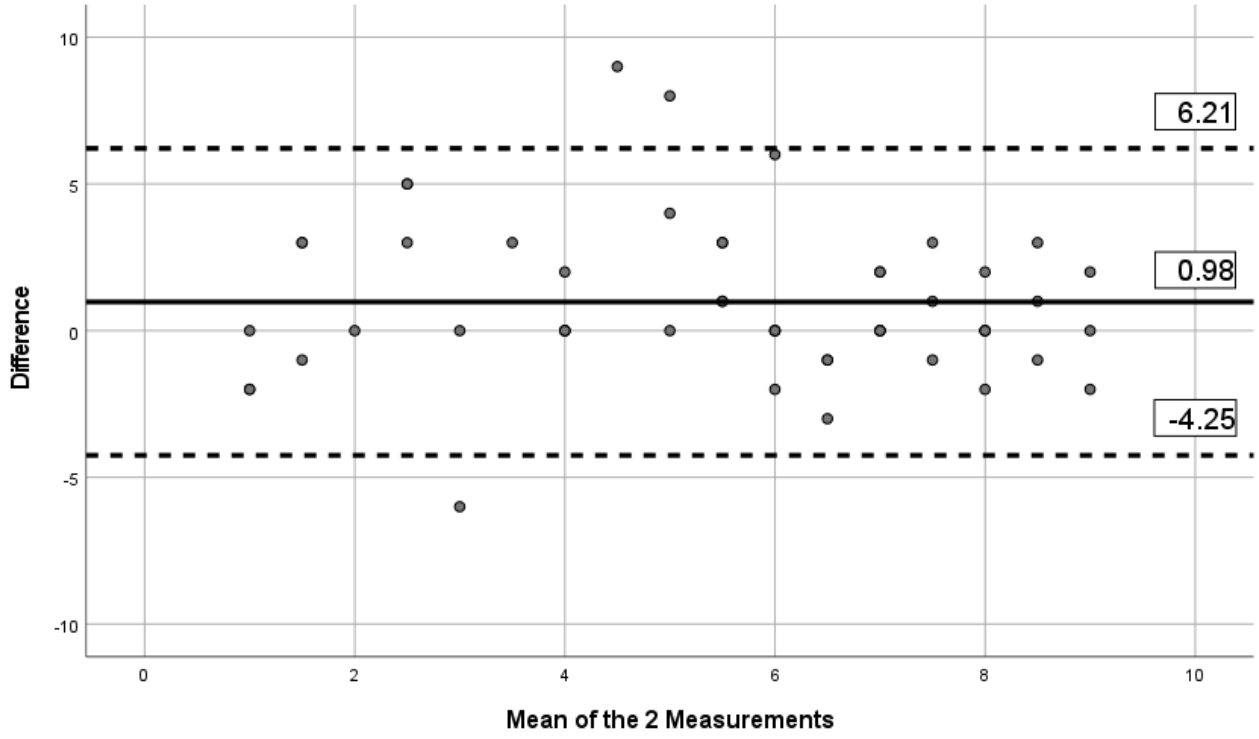


Figure 7. Bland-Altman limits for the number of painful body sites between the 2 assessments using the app.

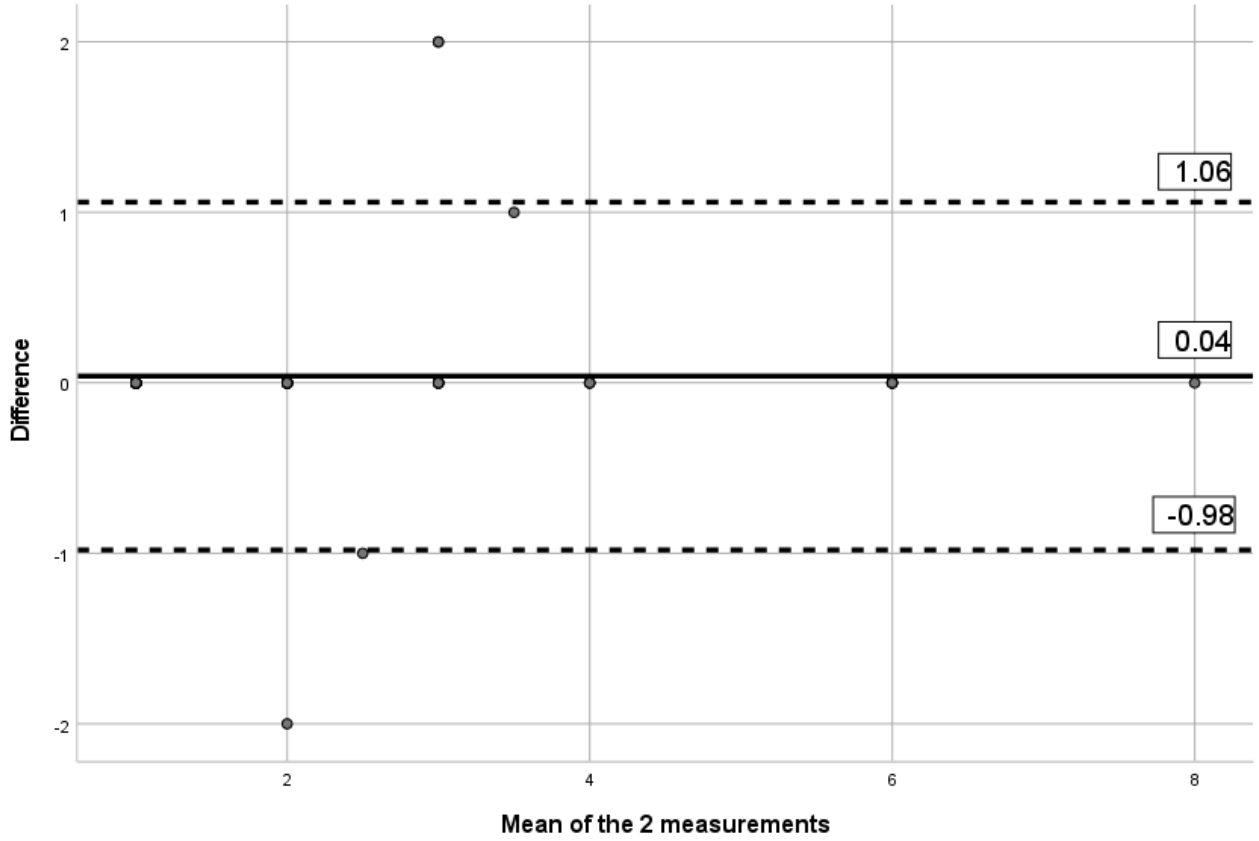


Table 6. Test-retest reliability and standard error of measurement of the AvaliaDor app.

Variable	Intraclass correlation coefficient (95% CI)	Standard error of measurement	Minimal detectable difference
Pain severity	0.86 (0.76-0.92)	1.96	5.43
Pain interference	0.84 (0.71-0.91)	6.40	17.73
Pain phenotype	0.90 (0.82-0.95)	2.35	6.52
Pain catastrophizing	0.76 (0.57-0.88)	1.30	3.60
Fear of movement	0.67 (0.42-0.81)	1.79	4.97
Number of painful body sites	0.98 (0.97-0.99)	0.24	0.66

Criterion Validity Assessment

The correlation between the paper version of the questionnaire and the app version was 0.84 for pain severity, 0.80 for pain interference, 0.84 for pain phenotype, and 0.93 for the number of painful body sites. Besides, visual inspection of the Bland-Altman limits (Figures 8-11) shows a symmetrical distribution around the mean, close to 0, with no systematic or proportional bias.

Figure 8. Bland-Altman limits for the Brief Pain Inventory pain interference subscale using the paper questionnaire and the mobile app.

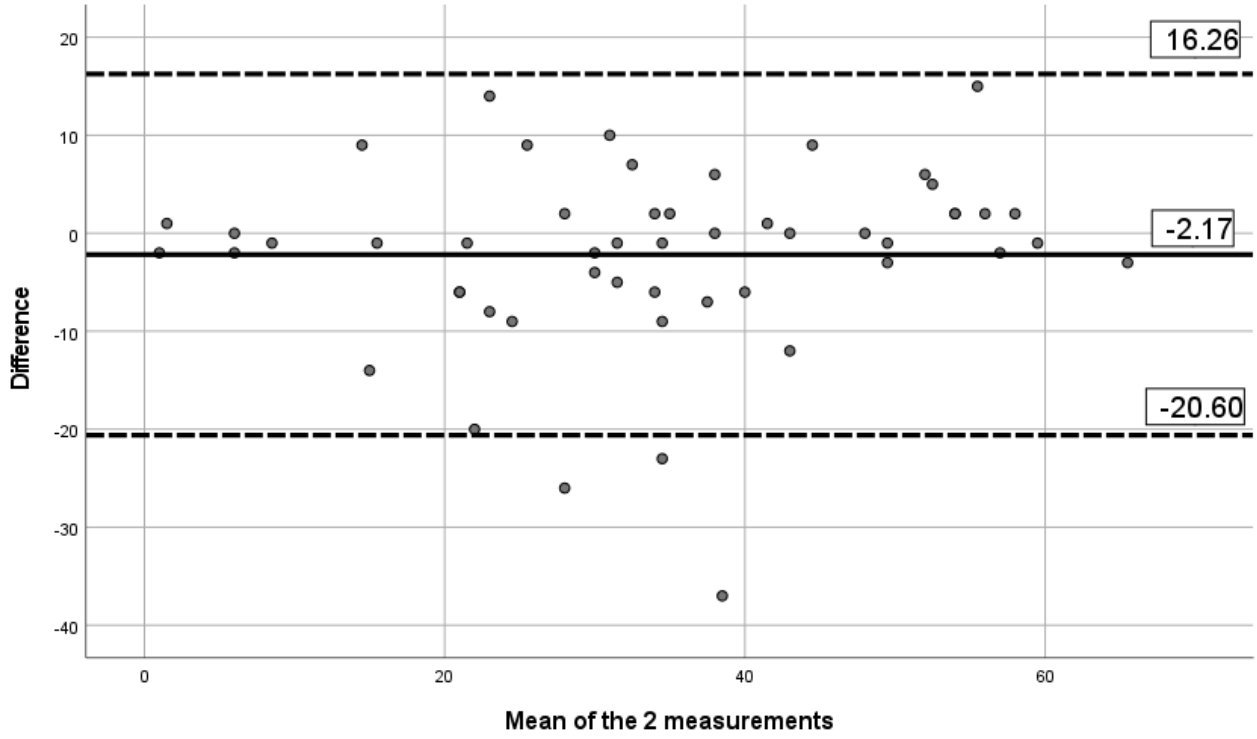


Figure 9. Bland-Altman limits for the Brief Pain Inventory pain severity subscale using the paper questionnaire and the mobile app.

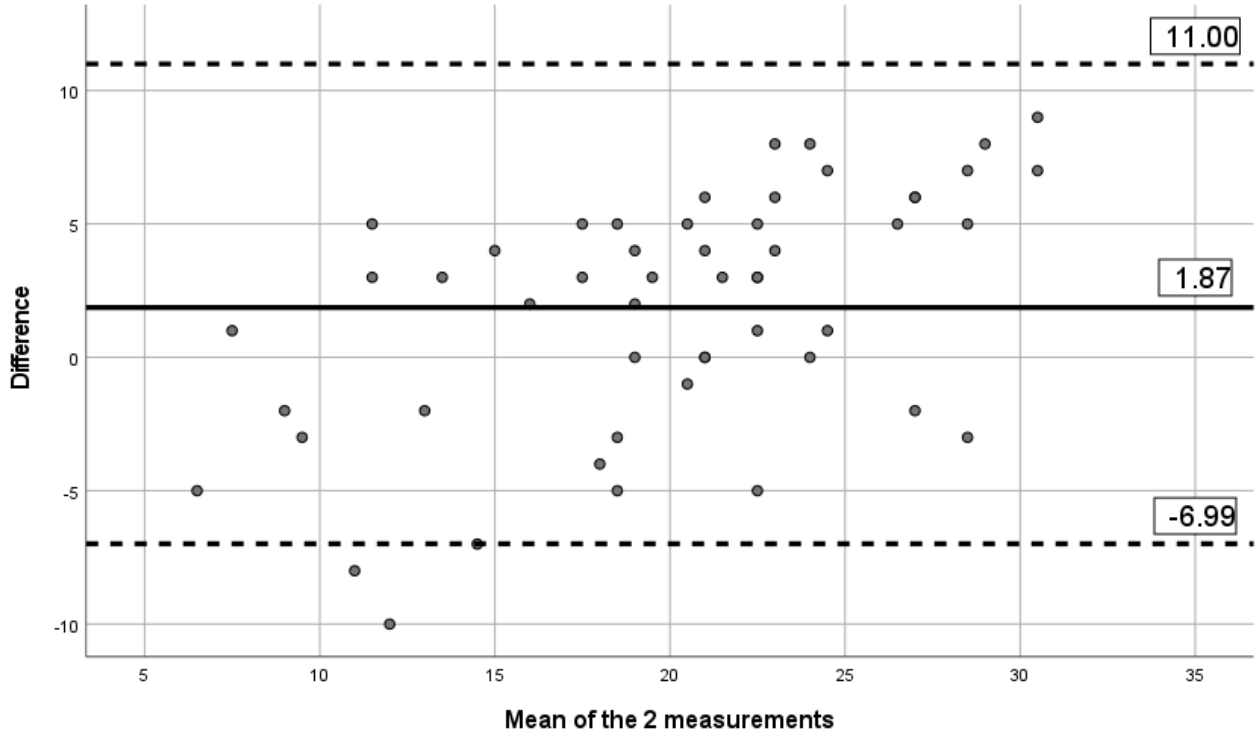


Figure 10. Bland-Altman limits for painDETECT using the paper questionnaire and the mobile app.

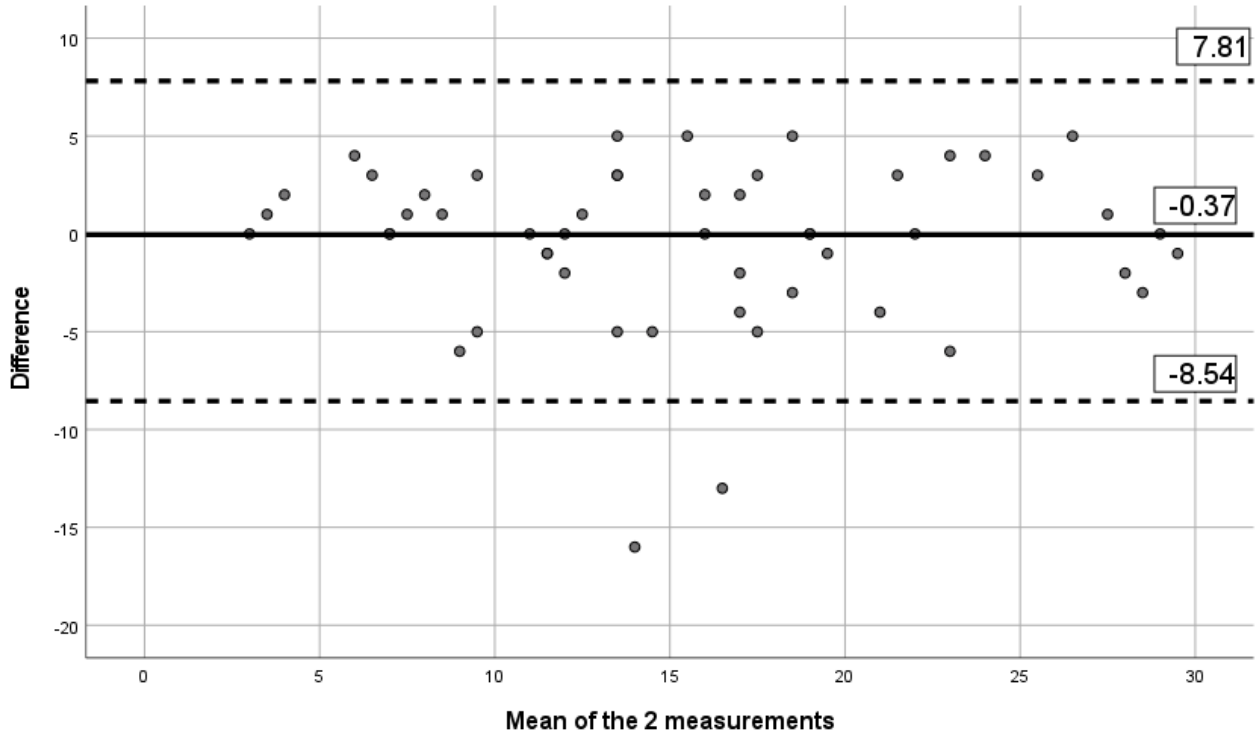
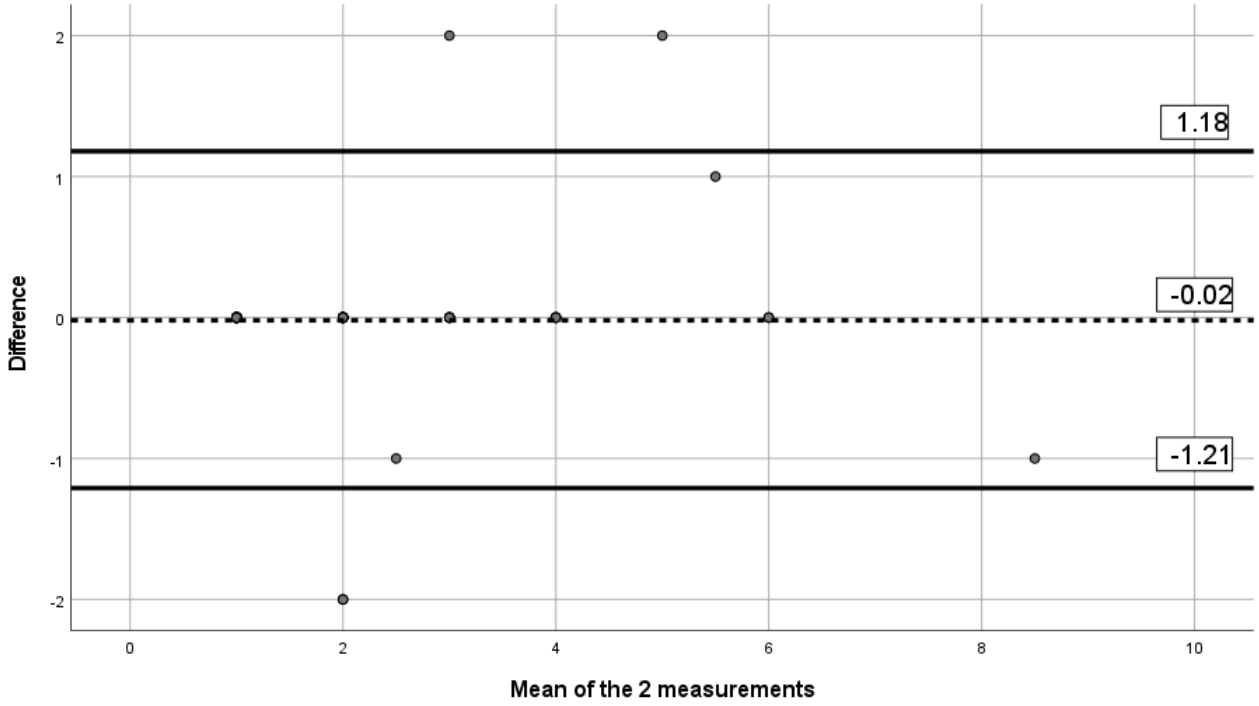


Figure 11. Bland-Altman limits for the number of painful body sites using the paper questionnaire and the mobile app.



Hypothesis Testing Assessment

The correlation between the Pain Catastrophizing Scale administered on paper and the single question on catastrophizing included in the app was r=0.59 (P<.001), indicating a strong correlation. The correlation between the Tampa Scale for Kinesiophobia and the single question on fear of movement included in the app was r=0.41 (P=.003), indicating a moderate correlation.

Discussion

Principal Findings

This paper presents the process of development of an app that aims to assess pain from a biopsychosocial perspective. The app assesses the intensity, location, and phenotype of pain; the associated disability; and the issues of pain catastrophizing and fear of movement. It was developed in close collaboration with patients and physiotherapists from a user-centered perspective. The results of 2 assessments in 2 different sessions were promising and suggest that the developed app (AvaliaDor) is valid, reliable, and usable.

A previous review reported that of 283 pain-related apps that were available in the main shops (eg, App Store and Google Play), none has undergone a scientific process of validation [13]. This finding undermines the trust that both patients and health professionals have in these solutions, compromising the potential and added value that these solutions can bring to the care of patients with pain. For example, mobile apps can capture real-time data (reducing the impact of memory bias on the outcomes), help clinicians reach a larger number of patients [13], and help detect a deterioration in the pain condition that may alert the clinician to schedule an appointment. However, the development of mobile apps through a user-centered and scientifically sound process is of utmost importance for the clinician to trust its results.

The development process of our app was based on guidelines of which aspects of pain should be assessed [22] but also on previous apps that were considered good examples and on the feedback of physiotherapists who were experts on pain assessment and treatment. This guaranteed that the app covered the relevant aspects of pain that need to be assessed to inform the treatment of patients with pain and to assess the evolution of these patients with treatment (ie, guarantees content validity). Also, the high correlation values between the paper versions of the questions for pain severity, pain impact, pain location, and pain phenotype and the same measurements using the AvaliaDor app shows that the app has criterion validity (ie, that the measures taken with the app are an adequate reflection of the "gold standard" paper-based questionnaires). The Bland-Altman plots supported these findings with small mean differences between the 2 methods of measurement (app versus paper-based questionnaires). The correlations between 1 question on pain catastrophizing and 1 question on fear of movement with the full paper-based Pain Catastrophizing Scale and the Tampa Scale for Kinesiophobia, respectively, were in line with the predefined hypothesis and suggest that the questions used in the app were measuring the respective constructs. The correlation values found for validity compare well with those reported in previous studies. For example, the developers of the Keele Pain Recorder app reported correlations of 0.79 for pain intensity and 0.60 for disability measured by the app through the question, "What is the interference of pain in the activities of the home, leisure or work," and on paper with the 36-item Short-Form Health Survey [23]. Another app, Pain Monitor, was developed to assess pain intensity, disability, and catastrophizing and used the same instruments in its validation process as were used in this study. The authors reported a strong

correlation for both pain intensity (r=0.71) and pain interference (r=0.67) when comparing the results obtained with the app and those obtained with the paper version of the Brief Pain Inventory [24].

In terms of test-retest reliability, the AvaliaDor mobile app presented excellent reliability for pain location (ICC 0.98); good reliability for pain phenotype (ICC 0.90), pain intensity (ICC 0.86), disability (ICC 0.84), and pain catastrophizing (ICC 0.76); and moderate reliability for fear of movement (ICC 0.65). These reliability values were identical to those of the paper questionnaires that served as the basis for the app questions [32,36]. However, the values of the SEM and MDD were relatively high, except those corresponding to the number of painful body sites. A difference between 2 measurements can only be considered to represent a true change in the patient's condition if it is larger than the MDD [16]. These findings suggest that the ability to detect small changes in the patient's condition over time may be affected. This may reflect the characteristic oscillatory behavior of chronic pain over time [45].

The mobile app AvaliaDor was considered to have good usability, which we believe is related to having taken a user-centered approach during its development [46]. The process of development included both physiotherapists and patients with pain from the very beginning—taking into account the users' needs and expectations regarding what the app should assess and how it should look and work (ie, user requirements)—in an interactive process that ended only when users were satisfied with the app. In addition, the development of the app was carried out by a multidisciplinary team comprised of 1 individual with expertise in information technologies and 2 musculoskeletal physiotherapists, thereby involving pain specialists from the conception phase of the app until its final version was produced. Usability assessment employed a mixed methods approach using indicators of performance (eg, time used to complete a full pain assessment using the app), interviews, and validated questionnaires, in line with recommendations [47]. The lack of inclusion of users and health professionals in the development process of digital solutions targeting the health care field is one of the limitations cited by several systematic reviews [14,48,49].

Study Limitations and Strengths

This study has limitations that must be considered. The data reported in this study were from 2 consecutive assessments only and further research is needed to assess the mobile app during prolonged use in a real-world context. Another limitation is that the mobile app is only available for the Android operating system. The strengths of the study are related to the use of the biopsychosocial model of pain assessment during the design of the app, the involvement of potential users from the beginning of the app development process, and the robust assessment of the app's usability, reliability, and validity.

Future Work

Future work can explore the long-term use of the AvaliaDor app in routine clinical pain management contexts and for different clinical pain conditions (eg, low back pain, neck pain, fibromyalgia), as well as the sensitivity of the app to detect changes due to interventions. Also, this mobile app can be used to facilitate high volumes of data collection, which can be analyzed with data analytics to explore potential data patterns that can inform the assessment and management of patients with pain.

Conclusions

A mobile app—AvaliaDor—was developed to assess the intensity, location, and phenotype of pain; the associated disability; and the issues of pain catastrophizing and fear of movement. It was developed in close collaboration with patients and physiotherapists from a user-centered perspective and was shown to be usable, valid, and reliable to assess pain from a biopsychosocial perspective in a heterogeneous group of patients with pain. Future work can explore the long-term use of AvaliaDor in clinical contexts and its advantages for the assessment and management of patients with pain.

Acknowledgments

This article was supported by National Funds through Fundação para a Ciência e a Tecnologia (FCT) within CINTESIS, R&D Unit (reference UIDB/4255/2020) and within IEETA/UA R&D Unit (UIDB/00127/2020).

Conflicts of Interest

None declared.

References

1. Schwab T, Langell J. Human Factors-Based Mobile Application Design for Global Health. Surg Innov 2018 Nov 22:1553350618814643. [doi: 10.1177/1553350618814643] [Medline: 30465467]
2. 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]
3. de Jong JM, Ogink PA, van Bunningen CG, Driessen RJ, Engelen LJ, Heeren B, et al. A Cloud-Based Virtual Outpatient Clinic for Patient-Centered Care: Proof-of-Concept Study. J Med Internet Res 2018 Sep 24;20(9):e10135 [FREE Full text] [doi: 10.2196/10135] [Medline: 30249584]
4. Global Observatory for eHealth. Global Diffusion of EHealth: Making Universal Health Coverage Achievable. In: Report of the third global survey on eHealth. Geneva: World Health Organization; 2016.

5. Global Observatory for eHealth. mHealth: New horizons for health through mobile technologie. In: Based on the findings of the second global survey on eHealth. Geneva: World Health Organization; 2011.
6. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
7. Vardeh D, Edwards R, Jamison R. There's an App for That: Mobile Technology Is a New Advantage in Managing Chronic Pain. Pain Clin Updates 2013;XXI(6):1-7.
8. Silva A, Queirós A, Caravau H, Ferreira A, Rocha N. Systematic Review and Evaluation of Pain-Related Mobile Applications. In: Cruz-Cunha MM, Miranda IM, Martinho R, Rijo R, editors. Encyclopedia of E-Health and Telemedicine. Pennsylvania: IGI Global; 2016:383-400.
9. Liaw S, Humphreys JS. Rural eHealth paradox: it's not just geography!. Aust J Rural Health 2006 Jun;14(3):95-98. [doi: 10.1111/j.1440-1584.2006.00786.x] [Medline: 16706876]
10. Talboom-Kamp EP, Verdijk NA, Harmans LM, Numans ME, Chavannes NH. An eHealth Platform to Manage Chronic Disease in Primary Care: An Innovative Approach. Interact J Med Res 2016 Feb 09;5(1):e5 [FREE Full text] [doi: 10.2196/ijmr.4217] [Medline: 26860333]
11. Jamison RN, Mei A, Ross EL. Longitudinal trial of a smartphone pain application for chronic pain patients: Predictors of compliance and satisfaction. J Telemed Telecare 2018 Feb;24(2):93-100. [doi: 10.1177/1357633X16679049] [Medline: 27831496]
12. 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]
13. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
14. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
15. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: A systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
16. van Kampen DA, Willems WJ, van Beers LWAH, Castelein RM, Scholtes VAB, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013 Nov 14;8:40 [FREE Full text] [doi: 10.1186/1749-799X-8-40] [Medline: 24225254]
17. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
18. Sandhu H, Wilson K, Reed N, Mihailidis A. A Mobile Phone App for the Self-Management of Pediatric Concussion: Development and Usability Testing. JMIR Hum Factors 2019 May 31;6(2):e12135 [FREE Full text] [doi: 10.2196/12135] [Medline: 31152527]
19. ISO 9241-210:2019: Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. ISO. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2019-12-10]
20. Almeida AF, Rocha NP, Silva AG. Methodological Quality of Manuscripts Reporting on the Usability of Mobile Applications for Pain Assessment and Management: A Systematic Review. Int J Environ Res Public Health 2020 Jan 27;17(3):785 [FREE Full text] [doi: 10.3390/ijerph17030785] [Medline: 32012674]
21. Streiner D, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 4th edition. Oxford: Oxford University Press; 2008.
22. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain 2016 Sep;157(9):1851-1871 [FREE Full text] [doi: 10.1097/j.pain.0000000000000602] [Medline: 27152687]
23. Bedson J, Hill J, White D, Chen Y, Wathall S, Dent S, et al. Development and validation of a pain monitoring app for patients with musculoskeletal conditions (The Keele pain recorder feasibility study). BMC Med Inform Decis Mak 2019 Jan 25;19(1):24 [FREE Full text] [doi: 10.1186/s12911-019-0741-z] [Medline: 30683106]
24. Suso-Ribera C, Castilla D, Zaragozá I, Ribera-Canudas MV, Botella C, García-Palacios A. Validity, Reliability, Feasibility, and Usefulness of Pain Monitor: A Multidimensional Smartphone App for Daily Monitoring of Adults With Heterogenous Chronic Pain. Clin J Pain 2018 Oct;34(10):900-908. [doi: 10.1097/AJP.0000000000000618] [Medline: 29659375]
25. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap 1994 Mar;23(2):129-138. [Medline: 8080219]
26. Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Further Validation of a Portuguese Version of the Brief Pain Inventory Interference Scale. Clín Salud 2012;23(1):89-96. [doi: 10.5093/cl2012a6]

27. Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006 Oct;22(10):1911-1920. [doi: 10.1185/030079906X132488] [Medline: 17022849]
28. Lutz J, Gross R, Long D, Cox S. Predicting Risk for Opioid Misuse in Chronic Pain with a Single-Item Measure of Catastrophic Thinking. J Am Board Fam Med 2017;30(6):828-831 [FREE Full text] [doi: 10.3122/jabfm.2017.06.170124] [Medline: 29180559]
29. Verwoerd AJH, Luijsterburg PAJ, Timman R, Koes BW, Verhagen AP. A single question was as predictive of outcome as the Tampa Scale for Kinesiophobia in people with sciatica: an observational study. J Physiother 2012;58(4):249-254 [FREE Full text] [doi: 10.1016/S1836-9553(12)70126-1] [Medline: 23177227]
30. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
31. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998 Jan 15;17(1):101-110. [doi: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e] [Medline: 9463853]
32. Lopes J, Patto T, Serra S. Questionários sobre Dor Crónica. Dor 2007;15(4):38-39.
33. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
34. Parkerson HA, Noel M, Pagé MG, Fuss S, Katz J, Asmundson GJG. Factorial validity of the English-language version of the Pain Catastrophizing Scale--child version. J Pain 2013 Nov;14(11):1383-1389. [doi: 10.1016/j.jpain.2013.06.004] [Medline: 24012382]
35. Sullivan MJL, Adams H, Sullivan ME. Communicative dimensions of pain catastrophizing: social cueing effects on pain behaviour and coping. Pain 2004 Feb;107(3):220-226. [doi: 10.1016/j.pain.2003.11.003] [Medline: 14736584]
36. Cordeiro N, Pezarat-Correia P, Gil J, Cabri J. Portuguese Language Version of the Tampa Scale for Kinesiophobia [13 Items]. J Musculoskelet Pain 2013 Apr 29;21(1):58-63. [doi: 10.3109/10582452.2012.762966]
37. Poquet N, Lin C. The Brief Pain Inventory (BPI). J Physiother 2016 Jan;62(1):52 [FREE Full text] [doi: 10.1016/j.jphys.2015.07.001] [Medline: 26303366]
38. Tampin B, Bohne T, Callan M, Kvia M, Melsom Myhre A, Neoh EC, et al. Reliability of the English version of the painDETECT questionnaire. Curr Med Res Opin 2017 Apr;33(4):741-748. [doi: 10.1080/03007995.2017.1278682] [Medline: 28161985]
39. Rosa A, Martins A, Costa V, Queirós A, Silva A, Rocha N. European Portuguese validation of the Post-Study System Usability Questionnaire (PSSUQ). 2015 Presented at: 10th Iberian Conference on Information Systems and Technologies, CISTI. Institute of Electrical and Electronics Engineers Inc; 2015; Aveiro p. 1-5. [doi: 10.1109/cisti.2015.7170431]
40. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018 May;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Lawrence Erlbaum Associates; 1998.
42. Book Review. Foundations of Clinical Research: Applications to Practice. Surv Ophthalmol 2002 Nov;47(6):598. [doi: 10.1016/s0039-6257(02)00362-4]
43. Donoghue D, Physiotherapy ResearchOlder People (PROP) group, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med 2009 Apr;41(5):343-346 [FREE Full text] [doi: 10.2340/16501977-0337] [Medline: 19363567]
44. Altman D, Bland J. Measurement in Medicine: The Analysis of Method Comparison Studies. Journal of the Royal Statistical Society. Series D (The Statistician) 1983;32:307-317.
45. Fillingim RB, Loeser JD, Baron R, Edwards RR. Assessment of Chronic Pain: Domains, Methods, and Mechanisms. J Pain 2016 Sep;17(9 Suppl):T10-T20 [FREE Full text] [doi: 10.1016/j.jpain.2015.08.010] [Medline: 27586827]
46. User-Centered Design Basics. usability.gov. URL: https://www.usability.gov/what-and-why/user-centered-design.html [accessed 2020-01-10]
47. Silva AG, Simões P, Santos R, Queirós A, Rocha NP, Rodrigues M. A Scale to Assess the Methodological Quality of Studies Assessing Usability of Electronic Health Products and Services: Delphi Study Followed by Validity and Reliability Testing. J Med Internet Res 2019 Nov 15;21(11):e14829 [FREE Full text] [doi: 10.2196/14829] [Medline: 31730036]
48. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of Self-Management Support Functions in Apps for People With Persistent Pain: Systematic Review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
49. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]

Abbreviations

ICC: intraclass correlation coefficient
IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
MDD: minimal detectable difference
MDD95: minimal detectable difference with a confidence level of 95%
mHealth: mobile health
SEM: standard error of measurement

Edited by L Buis; submitted 27.10.20; peer-reviewed by T Alessa, I Schiering; comments to author 02.12.20; revised version received 21.12.20; accepted 08.03.21; published 14.05.21.

Please cite as:
Lopes F, Rodrigues M, Silva AG
User-Centered Development of a Mobile App for Biopsychosocial Pain Assessment in Adults: Usability, Reliability, and Validity Study
JMIR Mhealth Uhealth 2021;9(5):e25316
URL: https://mhealth.jmir.org/2021/5/e25316
doi:10.2196/25316
PMID:33988515

Original Paper

Behavior Change Techniques in Popular Mobile Apps for Smoking Cessation in France: Content Analysis

Luz Adriana Bustamante[1], MSci; Cédric Gill Ménard[1], MSci; Sabrina Julien[2], PhD, MCF; Lucia Romo[1], Pr, MCF

[1]Laboratoire EA 4430-Clinique Psychanalyse Developpement, Department of Psychology, University of Paris Nanterre, Nanterre, Ile-de-France, France
[2]C2S - Cognition Santé Société (EA 6291), Department of Psychology, Université de Reims Champagne - Ardenne, Reims, France

Corresponding Author:
Luz Adriana Bustamante, MSci
Laboratoire EA 4430-Clinique Psychanalyse Developpement
Department of Psychology
University of Paris Nanterre
200, Avenue de la Republique
Nanterre, Ile-de-France, 92000
France
Phone: 33 783192547
Email: adriana.bustamante93@gmail.com

Abstract

Background: The mobile app market differs from country to country, and to date, no previous review of the content quality of smoking cessation apps has been conducted in France.

Objective: This study aimed to examine the general quality of the most popular smoking cessation apps in France and also determine the degree to which apps adhere to established behavioral and cognitive techniques (BCTs) proven effective in clinical practice.

Methods: A systematic research of smoking cessation apps was conducted in both the Google Play Store and Apple Store in the French market. The general quality of popular apps was rated with the Mobile App Rating Scale (MARS), and the therapeutic quality was assessed with the ratio of adherence of the behavior change technique taxonomy for smoking cessation treatment.

Results: A total of 14 mobile apps met all the inclusion criteria of the content analysis. The interrater reliability varied from "substantial" (0.79) to "almost perfect" (0.9) for the two measures. The mean MARS score was 3.5 out of 5 (median 3.6, IQR 0.6 [3.2-3.8]). The findings suggest that popular apps focus primarily on the functionality dimension of the MARS scale (4.2/5). The mean number of BCTs was 22, with a large difference between apps (minimum 4, maximum 38). At least half of the apps addressed motivation (8.8/14, 63%) and advised on using behavioral skills in order to quit smoking or stay a nonsmoker (8.7/14, 62%). However, only a handful of apps gathered important information (5.9/14, 42%) in order to deliver proper advice regarding the use of approved medication or the implementation of behavioral techniques (4.3/14, 31%). The mean MARS score was positively correlated with the price (r=0.70, P=.007) and the number of BCTs used (r=0.67, P=.01). User rating was not correlated with any quality scale (P=.67).

Conclusions: The content quality of popular smoking cessation apps in France varied by app type and price. Most popular apps propose in general good quality content but lack implementation of evidence-based BCTs associated with effectiveness on smoking cessation treatment. Further research is needed to evaluate the improvement in the content quality of smoking cessation apps in France.

(JMIR Mhealth Uhealth 2021;9(5):e26082) doi:10.2196/26082

KEYWORDS

smartphone app; smoking cessation; mHealth; app quality; user engagement; behavior change technique taxonomy

Introduction

Despite a significant decrease in tobacco consumption in France from 30% in 2000 to 25.4% in 2018 and implementation of corresponding antismoking policies [1] and proven effective treatments [2], the prevalence of smokers aged 18 to 75 years is still a public health issue [1].

Among the new solutions proposed, mobile apps appear to be a promising form of support [3,4]. By adapting and transposing therapeutic principles already proven to be effective, apps may offer multiple benefits for patients, health care professionals, and the health care system itself [5-7]. One of the distinct mobile app benefits is ease of access to health care and therapeutic information [6]. In addition, an app-based digital health approach would help individuals foster a sense of responsibility and commitment to their own personal health through "nudges" like positive reinforcement via messaging, tracking of habits and regular feedback, and audiovisual support [5,7,8].

Although research on mobile apps is growing, it still lags behind innovation and business expansion [9]. The number of available smoking cessation apps is growing, as is their user base. For example, in 2009, there were 62 such apps available in the US market, and 3 years later, the number had quadrupled to 252 [10,11]. The number of apps has surely grown since then and far surpasses the number of peer-reviewed studies conducted in the same time period. As a Cochrane review notes, despite this proliferation of smoking cessation apps, there is insufficient evidence to conclude that they have a significant positive effect on long-term smoking cessation [12], even if studies conducted specifically on text message interventions have shown efficacy in increasing smoking cessation rates by 50% to 60% [12].

Beyond a relative dearth of studies, one factor that may explain the uncertainty surrounding the efficacy of these apps is that many do not integrate therapeutic approaches that have already been proven effective in other contexts. Indeed, all apps studied in the Cochrane review used different cognitive and behavioral theories as their point of departure [12,13]. The "active ingredients" (a term used to encapsulate the various strategies and practices of evidence-based behavioral and cognitive therapies) of strategies that have withstood clinical study and peer review and that are often recommended by public health authorities for clinical practice are not well integrated into the various apps available, whether in the American [11,14], Australian [15], or British [16] market.

Moreover, even if an app includes all the therapeutic guidelines, it might not be used; thus, its effectiveness would be limited. As Nielsen points out, 25% of the most downloaded apps are never used and 38% of purchased apps are immediately uninstalled after their first use [17]. The quality of user experience can also impact an app's efficacy. Undoubtedly, "adherence to therapies is a primary determinant of treatment success" [18], and health apps are unlikely to be an exception. Therefore, it is essential to identify and assess the main factors underlying the quality of user experience, defined by O'Brien and Toms as "user engagement with technology" [19]. To date, several factors that interfere with the quality of user experience have been identified, and alternative forms of assessment have been proposed [20]. For example, the Mobile App Rating Scale (MARS) proposes the following four quality dimensions: engagement, functionality, esthetics, and information [20]. This scale has been used to assess the quality of various mobile health (mHealth) apps in diverse health fields from weight management [21] and drug interaction [22] to smoking cessation [15].

The mobile app market differs from country to country, and to date, no previous review of smoking cessation app content has been conducted in France. This means the findings of previous reviews on English-language apps may not be transposable to a French context. Besides, most studies do not take into account both the user experience and the therapeutic aspect to assess the quality of existing mobile smoking cessation apps. This review aimed to use the MARS to examine the general quality of the most popular iOS and Android smoking cessation apps and use the behavior change technique taxonomy to determine the adherence degree to established behavioral and cognitive techniques (BCTs) proven effective in clinical practice [23].

Methods

Study Design

The apps analyzed in this study were searched for and downloaded in France using both the iOS and Android app stores. The names of all apps and their descriptions were initially screened by the first author (LAB). Most of the apps were downloaded for a second screening, and only those that met all criteria were recorded on video as if the user had downloaded them for the first time. Based on these videos, two independent raters assessed the presence of BCTs and mobile app quality. The raters were all trained in health behavior change, and they are both behavioral and cognitive psychologists and researchers. Since subjects were not recruited, no ethics approval was required. The recorded videos were necessary to ensure that both raters assessed the same versions of the apps and to facilitate discussion.

Sample

The study included both free and paid apps that support the French language and that intend to assist users with smoking cessation. The apps were identified on two occasions by the first author (LAB) (December 1, 2019, and April 20, 2020), using the app search keywords "smoking cessation," "stop smoking," and "quitting tobacco" (in French, "sevrage tabagique," "arrêter de fumer," and "arrêt tabac," respectively). To be included in the full review, apps had to be designed to target smoking cessation only and support the French language. Excluded apps were those that were last updated before January 2019 and had ratings of less than 3/5 points in the app stores. Since each store provides different information on the number of downloads, we used different characteristics to select the most popular apps for a detailed content analysis. For android apps, where the number of downloads is available, an arbitrary threshold of 50,000 downloads was required to meet our popularity criteria, and since Apple Store does not share the number of app downloads, for iOS apps, an arbitrary threshold of 50 raters per app was set. Smokers wanting to quit with an app seem to favor apps with a strong "social proof," meaning they care about the rating as well as the number of raters and number of downloads [24].

Procedure: Coding of Apps

After the first author used and recorded each app, both raters assessed the apps' content independently using both the MARS and behavior change technique taxonomy for smoking cessation.

Prior to evaluation, all raters read each type of measure and had the opportunity to clarify and discuss the definitions in order to ensure clear differentiation between items.

MARS

The MARS is a multidimensional measure for classifying and assessing the quality of mobile apps. The MARS total score can be used to evaluate and compare the quality of an app with others, while the subscale "subjective quality" can be used to describe the strengths and weaknesses of a specific app. The total score is calculated by averaging the mean scores of the following five categories: user engagement, functionality, esthetics, information, and subjective quality. Each category is rated using a 5-point scale ranging from inadequate (1) to excellent (5) [20]. This scale has already been used for assessing the quality of smoking cessation apps in the Australian market, with a high interrater reliability (IRR) between raters (interclass correlation coefficient [ICC]=0.807) [15]. The scale is largely used and translated in different languages with a high intraclass correlation coefficient and good internal consistency [25,26].

Behavior Change Technique Taxonomy

The behavior change technique taxonomy for individual behavioral support for smoking cessation was used in this study [27]. A dichotomous score of "0" (absent) or "1" (present) was applied for each technique during the assessment of every app [21,27]. Each technique was classified within the following functions that are needed to ensure the efficacy of cognitive and behavioral therapy for smoking cessation: (1) focus on behavior (B), (2) addressing motivation (M), (3) maximizing self-regulatory skills (S), (4) promoting adjuvant activities (A), (5) general aspects of interaction (R), (6) information gathering (I), (7) general communication (C), and (8) delivery of the intervention (D) [27].

Statistical Analysis

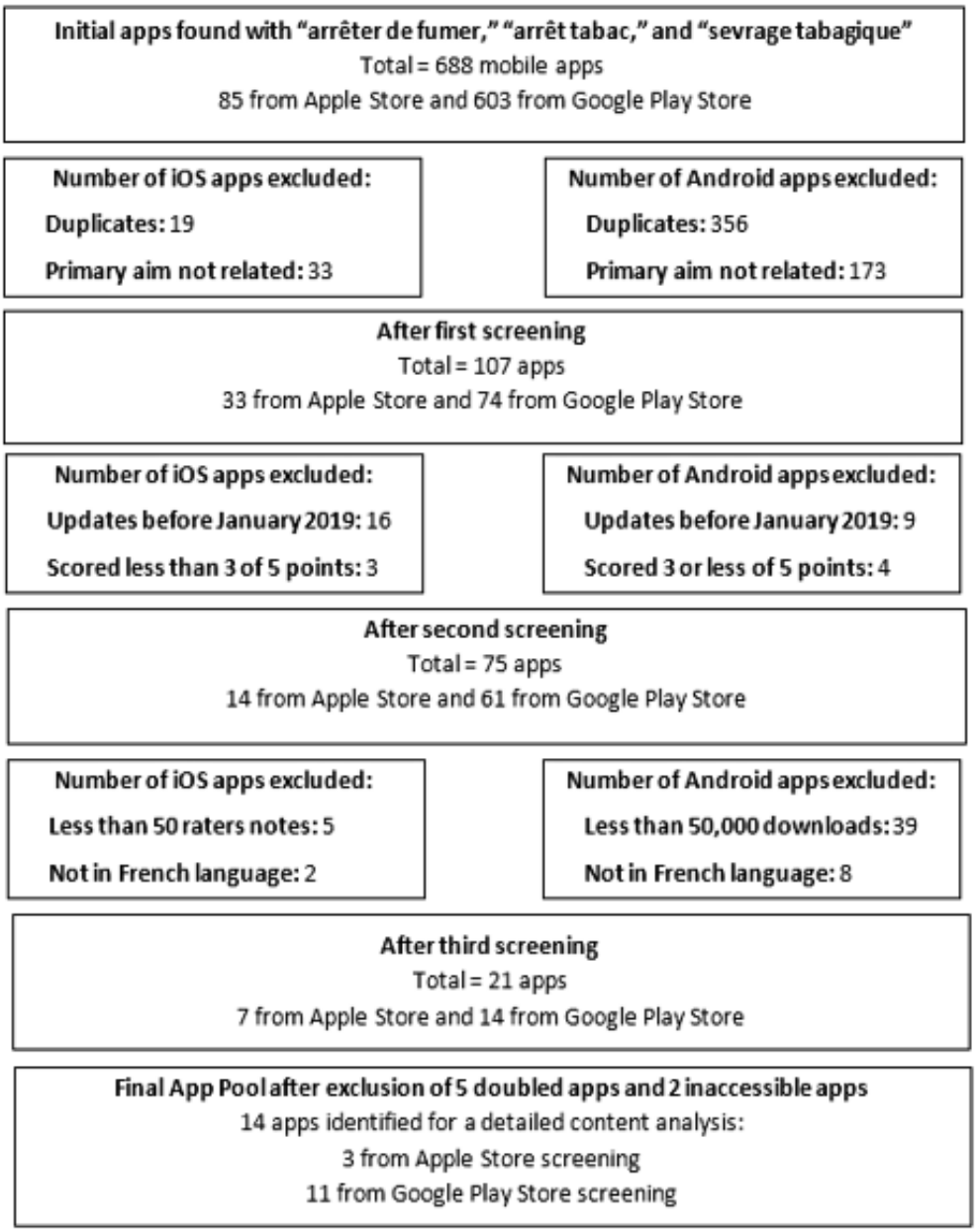
The statistical analysis was performed using IBM SPSS statistics version 26.0 (IBM Corp). Following the suggestion of "issues and best practices in content analysis" [28], we decided to calculate three measures of reliability for each scale. To assess the level of agreement between raters (IRR), we used the ICC and Krippendorff alpha for both scales, weighted kappa for the MARS, and prevalence and adjusted kappa for the behavior change technique taxonomy. The ICC was assessed in a two-way random model for an agreement level. The weighted kappa was assessed by pulling quadratic weights for each value. All reliability tests were performed per dimension of each scale and for all apps. Descriptive analysis was used to identify the presence of app characteristics (ie, mean price and frequency of BCTs), and one-way analysis of variance (ANOVA) and the post-hoc Tukey honestly significant difference (HSD) test were used to determine any significant difference observed between the scales. The mean score by dimension of each scale was used in the Spearman correlation test to examine the relationship among the price per month, user ratings, and both mobile app qualities (general and therapeutic).

Results

Systematic Search Results

A total of 688 apps were initially identified from the French Google Play Store (n=603) and Apple Store (n=85). Figure 1 shows the results of the key stages of the mobile app review. After preliminary inclusion and exclusion criteria were applied, 74 Android and 33 iOS apps remained. Further screening based on evaluation of the app product page and the last update resulted in 61 Android and 14 iOS apps. Finally, further exclusion of apps upon download and during the coding procedure resulted in a total of 14 apps. Among these 14 apps, seven were accessible in both stores, six were accessible in the Google Play Store only, and one was accessible in the Apple Store only (Multimedia Appendix 1).

Figure 1. Flowchart of the results from the app search, preliminary inclusion and exclusion screening, and final app pool.



General App Characteristics

Table 1 shows that most of the included apps (12/14, 86%) were affiliated with a commercial company, while 14% (n=2) were affiliated with a university (n=1) and a government department (n=1). The mean user rating was very good (mean 4.5, range 3.7-4.8) in both stores without a particular difference. All apps were free to download and to use for a limited time or with limited features, but half of them had in-app purchases with different payment options that offered full access to all the contents of the app. The mean monthly price was €3.47 (US$ 4.19) in both stores, without a difference between both stores. Most apps included calculator (14/14, 100%), tracker (13/14, 93%), and information (13/14, 93%) features, as well as media connectors and reminders to use the app or to stay a nonsmoker. Very few apps (4/14, 29%) allowed users to protect their information with a password. All the apps needed internet access to download, and later, most of the features could be used offline.

Table 1. General information of the rated apps from the French Apple Store and Google Play Store (N=14).

Variable	Total (n=14)	iOS (Apple Store) (n=8)	Android (Google Play Store) (n=13)
Affiliation, n (%)			
Commercial	12 (86%)	6 (75%)	11 (85%)
Unknown	0 (0%)	0 (0%)	0 (0%)
Government	1 (7%)	1 (13%)	1 (8%)
NGO[a]	0 (0%)	0 (0%)	0 (0%)
University	1 (7%)	1 (13%)	1 (8%)
Country of origin, n (%)			
England	1 (7%)	1 (13%)	1 (8%)
Spain	1 (7%)	1 (13%)	1 (8%)
France	4 (29%)	4 (50%)	1 (8%)
Switzerland	1 (7%)	1 (13%)	1 (8%)
Germany	4 (29%)	1 (13%)	4 (31%)
Ukraine	2 (14%)	0 (0%)	2 (15%)
United States	1 (7%)	0 (0%)	1 (8%)
Price			
Free, n (%)	6 (43%)	3 (38%)	6 (46%)
In-app payment, n (%)	8 (57%)	3 (38%)	7 (54%)
Price per month (€), mean (range; SD)	3.03 (0-9.99; 3.4)	3.68 (0-9.99; 3.5)	3.03 (0-9.99; 3.4)
Technical aspects, (%)			
Allows sharing	14 (100%)	8 (100%)	13 (100%)
Community	9 (64%)	7 (88%)	9 (69%)
Password protection	4 (29%)	3 (38%)	4 (31%)
Requires login	10 (71%)	6 (75%)	10 (77%)
Sends reminders	14 (100%)	8 (100%)	13 (100%)
Web access function	0 (0%)	0 (0%)	0 (0%)
Specific features, n (%)			
Calculator	14 (100%)	8 (100%)	13 (100%)
Rationing	2 (14%)	0 (0%)	1 (8%)
Tracker	13 (93%)	7 (88%)	12 (92%)
Information	13 (93%)	7 (88%)	13 (100%)
Game	6 (43%)	4 (50%)	6 (46%)
Lung health monitor	0 (0%)	0 (0%)	0 (0%)
Other	6 (43%)	5 (63%)	6 (46%)
Popularity			
User rating, mean (range)	4.5 (3.7-4.8; 0.32)	4.4 (4.1-4.7; 0.21)	4.4 (3.7-4.8; 0.32)
Number of ratings, mean (range)	__[b]	16,031 (74-83,000)	21,701 (221-86,713)
Store			
Apple only	1 (7%)	N/A[c]	N/A
Google Play only	6 (43%)	N/A	N/A
Both stores	8 (57%)	N/A	N/A

[a]NGO: nongovernmental organization.

[b]Not possible to measure.

[c]N/A: not applicable.

General Quality: MARS

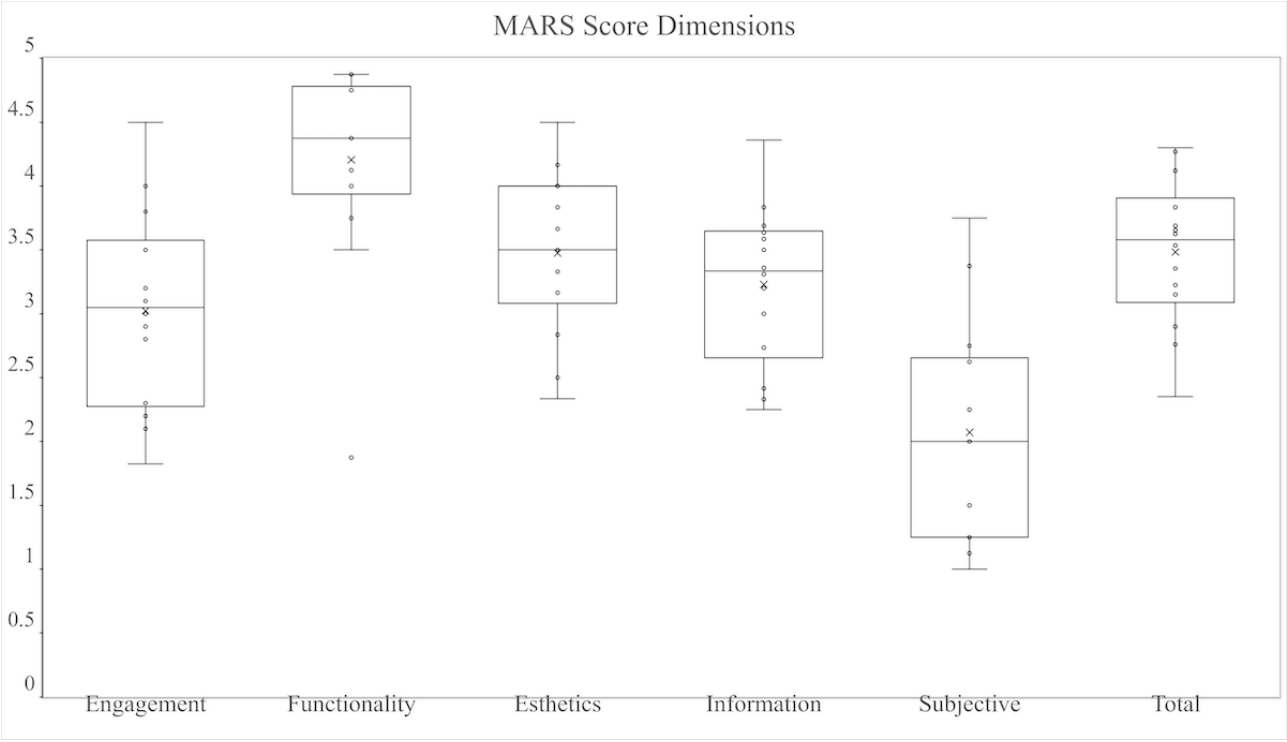
The general quality was acceptable. The mean MARS score was 3.5 (median 3.6, IQR 0.6 [3.2-3.8]), with a maximum score of 4.3 and a minimum score of 2.4 (Table 2). Since only one app was not available in both the Google Play Store and Apple Store, no comparison between these distribution services was conducted. The IRR between the two raters was substantial. The ICC was 0.79 (95% CI 0.74-0.84), weighted kappa was 0.79 (95% CI 0.74-0.84), and Krippendorf alpha was .88 (95% CI .85-.91). Detailed results are presented in Multimedia Appendix 2.

Table 2. Quality of smoking cessation mobile apps in the French market according to the Mobile App Rating Scale (MARS).

Mobile App Rating Scale (MARS) category	Mean (SD) score	Score range (minimum-maximum)	IQR (Q1-Q3)
Engagement (fun, interest, interactivity...)	3.0 (0.8)	(1.8-4.5)	(2.4-3.4)
Functionality (app functioning, easy to learn...)	4.2 (0.8)	(1.9-4.9)	(4.0-4.8)
Esthetics (overall visual appeal, stylistic consistency...)	3.5 (0.6)	(2.3-4.5)	(3.2-4.0)
Information (text, feedback, measures...)	3.2 (0.6)	(2.3-4.4)	(2.8-3.6)
Subjective (recommendation, overall rating...)	2.1 (0.9)	(1.0-3.8)	(1.3-2.6)
Total	3.5 (0.6)	(2.4-4.3)	(3.2-3.8)

The mean scores of the dimensions of the MARS are presented in Table 2. The values vary from “low” to “good.” The results of the one-way ANOVA reveal that there was at least one significant difference between the five dimensions regarding the score ($F_{5,78}$=13.51, P<.001). In order to determine which mean values differed more specifically from each other, the Tukey test (HSD) was performed. The results showed that the functionality dimension value was significantly higher than the values of the other dimensions (P=.009). The results also showed that the subjective dimension value was significantly lower than the values of all other dimensions (P=.001). However, the difference in the mean values of the engagement, information, esthetic, and total dimensions was not significant (P=.09). The engagement dimension had an average score but had the most variability (median 3.1, IQR 1.3). Figure 2 shows box plots with the median, first and third quartiles, and minimum and maximum scores. Each point represents the score for each app.

Figure 2. Mobile App Rating Scale (MARS) dimension scores.



Therapeutic Quality: Behavior Change Techniques

The mean number of BCTs identified in the apps was 22 (SD 10), with a maximum of 38 techniques and a minimum of 4 techniques. The IRR between the two raters was almost perfect. The ICC was 0.92 (95% CI 0.89-0.93), prevalence and adjusted kappa was 0.85 (95% CI 0.84-0.85), and Krippendorf alpha was .85 (95% CI .79-.9). Detailed results are presented in Multimedia Appendix 3.

The significance in the Shapiro-Wilk test (P=.11) indicated the inability to assess the strengths of the different prevalences of BCTs observed in each app. Yet, delivery of the intervention (D), addressing motivation to stay a nonsmoker (M), specific behavior change techniques focused on smoking behavior (B), and maximizing self-regulatory capacity and skills (S) were observed most frequently (between 62% [8.7/14] and 64% [9/14] of the apps), whereas proposing adjuvant strategies (A) was least frequently observed (31% [4.3/14] of the apps) (Table 3) [27].

Table 3. Behavioral and cognitive techniques classified by function in the smoking cessation apps studied, according to the taxonomy of Michi et al [27].

BCTs[a] classified by function	Mean (SD) app number (out of 14)	Range (minimum-maximum)	Percentage
Delivery of the intervention (D) (ie, provide adapted behavioral support)	9.0 (1.1)	(1.5-13)	64%
Addressing motivation to stay a nonsmoker (M) (ie, provide information of the advantages of quitting)	8.8 (3.5)	(1.5-12.5)	63%
Specific smoking behavior change techniques (B) (ie, Monitor how the client achieves his goal)	8.7 (3.0)	(1.5-13)	62%
Maximizing self-regulatory capacity and skills (S) (ie, provide advise on how to avoid social pressure)	8.6 (2.7)	(2.5-13)	62%
Information gathering (I) (ie, evaluate the patient's readiness to quit)	5.9 (2.2)	(3-7.5)	42%
General communication (C) (ie, provide information on withdrawal symptoms)	5.8 (3.3)	(0-10)	41%
General aspects of interaction (R) (ie, encourage or reassure on client experiences)	5.3 (3.1)	(0-10)	38%
Adjuvant activities (A) (ie, explain the advantages of medication if needed)	4.3 (4.1)	(0-8.5)	31%

[a]BCTs: behavioral and cognitive techniques.

Relationship Between App Characteristics and Quality Level

Mean price per month, mean user rating, and number of BCTs were tested for correlation with the MARS score (Table 4). The mean MARS score (mean 3.5, SD 0.6) was positively correlated with price (mean 3.0, SD 3.4) (r=0.70, P=.007) and the number of implemented BCTs (mean 22, SD 11) (r=0.66, P=.01). User rating was not correlated with any quality scale (Table 4). The mean price per month (mean 3.0, SD 3.4) was positively correlated with the mean user rating (mean 4.5, SD 0.3) (r=0.58, P=.03) and the number of BCTs in the app (mean 22, SD 11) (r=0.60, P=.03).

Table 4. Correlations among the total Mobile App Rating Scale (MARS) score, price, user rating, and number of behavioral and cognitive techniques.

Variable	Mean price (€)	User mean rating score	MARS[a] score	Total number of BCTs[b]
Mean price (€)				
r	1	0.58	0.70	0.59
P value	—[c]	.03	.007	.03
User mean rating score				
r	0.58	1	0.098	0.124
P value	.03	—	.74	.67
MARS score				
r	0.70	0.098	1	0.66
P value	.007	.74	—	.01
Total number of BCTs				
r	0.59	0.124	0.66	1
P value	.03	.67	.01	—

[a]MARS: Mobile App Rating Scale.

[b]BCTs: behavioral and cognitive technique.

[c]Not applicable.

Discussion

Systematic Search Results

The current review aimed to examine the content quality of popular mobile apps for smoking cessation in the French market. This type of content analysis is, to our knowledge, the first of its kind for the following two main reasons: the target market of the study and the methodology. French mHealth apps were reviewed in this study, with the aim to examine both of the following aspects of content quality: the general quality via the MARS scale and the therapeutic dimension though the behavior change technique taxonomy.

General App Characteristics

Based on the established behavior change technique taxonomy and the MARS, we analyzed a total of 14 apps. It appears that the French mobile app market is less developed than the English one, where the number of reviewed apps is much higher. Indeed, 252 apps were identified in the US market in 2013 [11] and 225 apps were identified in 2016 [14]. Similar to these findings, 112 apps were examined in Australia [15] and 140 apps in England [16]. Consequently, even if several apps are available on the market, they are not accessible to most of the native French-speaking population speaking only French. This accessibility limitation could be overcome by translating existing apps on the market. However, the translation process would need to take into account the cultural context and be periodically adapted to each update to ensure users can access and benefit from the app content. Since this process would require human time and financial resources, the translation of apps could be decided based on supply and demand.

General Quality: MARS

The general quality of popular apps in France varies from “acceptable” to “good.” In spite of these results, health professional judges do not recommend most of the apps. It seems the general quality threshold needed to be recommended is not met by most of the popular apps on the French market.

Our findings suggest that popular apps focused primarily on the functionality dimension that is composed of the following four aspects: performance, ease of use, navigation, and gestural design. The priority on the functionality dimension is a trend already observed for smoking cessation [29] and weight management [21] apps. Enhancing only this dimension will not be enough to improve the general quality of content. Nevertheless, we recognize the importance of the functionality dimension as a facilitator for the use of mHealth solutions. Better integration of clinical expertise seems to be necessary to create engaging and informative content.

Therapeutic Quality: Behavior Change Techniques

The use of BCTs in our mobile app sample was not normally distributed, indicating that evidence-based techniques are not properly implemented in the French market. Our findings show that the market is still in the initial stages, most likely driven by technical expertise, and suggest that there is a lack of theory-driven behavioral change techniques, as proposed previously [16,29]. In France, mobile app development is not driven by BCT theory because its techniques are rarely invoked in popular apps. Indeed, even the most widely used BCT functions were absent in more than 65% of the sample.

The strength of current apps is their focus on the target behavior through addressing motivation and maximizing self-regulatory capacity and skills. In contrast, our results suggest that most apps failed to include features from the adjuvant activities dimension and focus on the general aspect of interaction. Failure to promote activities that indirectly facilitate abstinence (ie, inform or advise on medication to stop smoking) is consistent with the tendency in English-speaking countries [10,30]. A lack of adjuvant activities results in harsh deprivation considering

that adherence to medications that help to stop smoking increases the likelihood of successful abstinence by 50% compared to "cold turkey" [2]. Additionally, our findings indicate a deficit of techniques necessary for effective delivery (ie, acquire and communicate relevant information needed to adapt the intervention). In addition, some authors pointed out the importance of technique interactions for an effective behavior change, promoting the idea that some BCTs can be effective solely under specific combinations (used at their best under specific conditions). As reported, the least effective interventions were those providing *feedback on performance* without *providing instructions* [31]. We therefore believe that the current strength of the French mHealth market can be severely undermined owing to the limited information and communication techniques used.

Relationship Between App Characteristics and Quality Level

A noteworthy result was the positive relationship between both general and therapeutic qualities. This association is relevant since efficacy is influenced by not only the content of a therapy (in this case the therapeutic quality), but also adherence to treatment [18].

The second interesting result was the absence of correlation between the app store's user rating and the qualities assessed, suggesting that user ratings may not be a good predictor of general or therapeutic quality. These results may give rise to concerns as users choose their apps based primarily on the ratings on the app store [32]. This highlights the importance of creating standards and accreditation for mHealth apps to protect users.

The third interesting result was the relation between price and both quality measurements (the cheaper the app, the poorer the qualities, with the exception of two apps financed by public institutions). This result is at odds with the supposed accessibility benefits of mobile apps [6], which casts a doubt on the benefits of affordability. To better understand this relationship, more studies that address the issue of intervention efficacy should address the issue of the cost-benefit ratio.

Strengths and Limitations

An important strength of this study is that it is the first to examine smoking cessation apps on both the Apple Store and Google Play Store in the French market. The IRR was evaluated following the best practice recommendations for content analysis [28], and in all the tests, the IRR varied from substantial to almost perfect for each scale.

The findings of this study should be interpreted in light of some limitations. First, the apps were rated from the first and only use. There is a probability that some BCTs were not accessible to the raters and therefore were underrated. On the contrary, after prolonged use, apps could be seen as less engaging than in the first use, as the engagement attrition rate of mHealth apps is very high [33,34]. Second, we are aware of critics questioning the link between the quality and popularity of smoking cessation apps (whether the quality is therapeutic or general). Indeed, the few apps that exhibited high adherence to therapeutic guidelines were not necessarily the most popular [16], and about 17% of the high general quality apps identified appeared in the top 10 recommended smoking cessation apps in New Zealand app stores [15]. The aim of our study was not to identify the best solution in the French market, but to establish the quality of the most used apps available for French users.

Conclusions

The general and therapeutic content quality of popular smoking cessation apps in France varied by app type and price. General and therapeutic contents are positively correlated. The user rating on app stores is not an indicator of the general and therapeutic content quality. The findings suggest that popular apps focused primarily on the functionality dimension. At least half of the apps addressed motivation and advised on using behavioral skills to quit smoking or stay a nonsmoker; however, only a handful of apps gathered important information and delivered proper advice regarding the use of approved medication or the implementation of behavioral techniques. Overall, the findings provide the first snapshot of the quality of popular smoking cessation apps in France. This review will need to be revised in order to examine whether the content quality of smoking cessation apps will evolve in the French market. Further research is needed to understand how users engage and benefit from these apps in the real world.

Acknowledgments

The research was supported and funded by the National Association of Research and Technology (ANRT) and Kwit SAS. The funders played no role in the design, conduct, or analysis of the study, or in the interpretation and reporting of the study's findings. The researchers and judges were independent from the funders.

Authors' Contributions

The idea for the study was conceived and the review design was established by LAB in 2019. The research and screening app selection were performed by LAB. LR and SJ independently rated the apps separately, using the two scales. LAB, LR, and CGM contributed to preparing and developing the manuscript. LAB and CGM generated the tables, figures, and flowchart, and organized the references. LAB reviewed and refined the manuscript.

Conflicts of Interest

LAB undertakes consultancy and research for a mobile app for smoking cessation (Kwit SAS) in a PhD contract involving the National Association of Research and Technology (ANRT). LR, CGM, and SJ have no conflicts to declare.

Multimedia Appendix 1
Detailed information of all apps included in the analysis.
[DOCX File , 16 KB - mhealth_v9i5e26082_app1.docx]

Multimedia Appendix 2
Interrater reliability with 95% CI according to each dimension of the Mobile App Rating Scale (MARS).
[DOCX File , 14 KB - mhealth_v9i5e26082_app2.docx]

Multimedia Appendix 3
Interrater reliability with 95% CI according to each dimension of the behavior change technique taxonomy.
[DOCX File , 14 KB - mhealth_v9i5e26082_app3.docx]

References

1. Andler R, Richard JB, Guignard R, Quatremère G, Verrier F, Gane J, et al. [Reduction of daily smoking rate among adults: results from the 2018 Santé publique France Health Barometer]. Bulletin Epidémiologique Hebdomadaire 2019;15:271-277 [FREE Full text]
2. Arrêt de la consommation de tabac : du dépistage individuel au maintien de l'abstinence en premier recours. Haute Autorité de Santé. 2014. URL: https://www.has-sante.fr/upload/docs/application/pdf/2014-01/recommandations_-_arret_de_la_consommation_de_tabac.pdf [accessed 2021-04-28]
3. mHealth: New horizons for health through mobile technologies: Based on the findings of the second global survey on eHealth. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-04-28]
4. WHO report on the global tobacco epidemic, 2017: Monitoring tobacco use and prevention policies. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255874/9789241512824-eng.pdf?sequence=1 [accessed 2021-04-28]
5. Green Paper on mobile health (mHealth). European Commission. 2014. URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2021-04-28]
6. Perroy AC. [m-health in an age of e-health. Promises, challenges and liabilities]. Ann Pharm Fr 2016 Nov;74(6):421-430. [doi: 10.1016/j.pharma.2016.03.002] [Medline: 27139163]
7. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
8. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
9. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
10. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
11. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
12. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
13. Regmi K, Kassim N, Ahmad N, Tuah N. Effectiveness of Mobile Apps for Smoking Cessation: A Review. Tob Prev Cessat 2017 Apr 12;3(Tobacco Prevention & Cessation):12-11 [FREE Full text] [doi: 10.18332/tpc/70088] [Medline: 32432186]
14. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How Smart are Smartphone Apps for Smoking Cessation? A Content Analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
15. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]

16. Rajani NB, Weth D, Mastellos N, Filippidis FT. Adherence of popular smoking cessation mobile applications to evidence-based guidelines. BMC Public Health 2019 Jun 13;19(1):743-748 [FREE Full text] [doi: 10.1186/s12889-019-7084-7] [Medline: 31196062]
17. Smartphones: So Many Apps, So Much Time. Nielsen. 2014. URL: https://www.nielsen.com/us/en/insights/article/2014/smartphones-so-many-apps--so-much-time [accessed 2020-02-10]
18. Adherence to long-term therapies: Evidence for action. World Health Organization. 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf [accessed 2021-04-28]
19. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
20. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
21. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
22. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
23. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
24. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25-14 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
25. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov S, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
26. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German Version of the Mobile App Rating Scale (MARS-G): Development and Validation Study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
27. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
28. Lacy S, Watson BR, Riffe D, Lovejoy J. Issues and Best Practices in Content Analysis. Journalism & Mass Communication Quarterly 2015 Sep 28;92(4):791-811. [doi: 10.1177/1077699015607338]
29. Ubhi HK, Kotz D, Michie S, van Schayck OCP, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
30. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
31. Dusseldorp E, van Genugten L, van Buuren S, Verheijden MW, van Empelen P. Combinations of techniques that effectively change health behavior: evidence from Meta-CART analysis. Health Psychol 2014 Dec;33(12):1530-1540. [doi: 10.1037/hea0000018] [Medline: 24274802]
32. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors Related to User Ratings and User Downloads of Mobile Apps for Maternal and Infant Health: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
33. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
34. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

Abbreviations

ANOVA: analysis of variance
BCT: behavioral and cognitive technique
HSD: honestly significant difference
ICC: interclass correlation coefficient
IRR: interrater reliability
MARS: Mobile App Rating Scale

mHealth: mobile health

Edited by L Buis; submitted 28.11.20; peer-reviewed by D Zhai, H Mehdizadeh; comments to author 08.01.21; revised version received 22.01.21; accepted 31.03.21; published 13.05.21.

Please cite as:
Bustamante LA, Gill Ménard C, Julien S, Romo L
Behavior Change Techniques in Popular Mobile Apps for Smoking Cessation in France: Content Analysis
JMIR Mhealth Uhealth 2021;9(5):e26082
URL: https://mhealth.jmir.org/2021/5/e26082
doi:10.2196/26082
PMID:33983130

Original Paper

Effective German and English Language mHealth Apps for Self-management of Bronchial Asthma in Children and Adolescents: Comparison Study

Julian Franzmair[1*], MD; Susanne C Diesner-Treiber[1*], MD, PhD; Julian J M Voitl[1], MD; Peter Voitl[1,2], MSc, MD

[1]First Vienna Pediatric Medical Center, Vienna, Austria

[2]Sigmund Freud University, Vienna, Austria

[*]these authors contributed equally

Corresponding Author:
Peter Voitl, MSc, MD
First Vienna Pediatric Medical Center
Donau-City-Strasse 1
Vienna, 1220
Austria
Phone: 43 1 263 7979
Fax: 43 1 263 7979
Email: peter.voitl@kinderarzt.at

Abstract

Background: Mobile health (mHealth) apps hold great potential for asthma self-management. Data on the suitability of asthma apps intended for children are insufficient, and the availability of German language apps is still inadequate compared with English language apps.

Objective: This study aims to identify functional asthma apps for children in German and to compare them with English language apps. In line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, the Google Play Store and Apple App Store are systematically searched to preselect the most efficient apps, which are then compared according to a self-compiled criteria catalog.

Methods: Both app stores were screened for the term *asthma*. Following a PRISMA preselection process, the apps that met the inclusion criteria (ie, available free of charge, German or English language, and suitable for children) were rated by 3 independent persons following a criteria catalog consisting of 9 categories, some conceived for this purpose (*availability*, *child-friendly*, *learning factor*, and *range of functions*) and some adopted from existing validated catalogs (*functionality and design*, *ease of use*, *potential for improving asthma self-management*, *fun factor and incentives*, and *information management and medical accuracy*). The highest rated apps in German and English were compared.

Results: A total of 403 apps were identified on the Google Play Store and the Apple App Store. Finally, 24 apps that met the inclusion criteria were analyzed. In the first step of the quality assessment, only 4 available German language asthma apps were compared with 20 English language asthma apps. The 4 German language apps were then compared with the 4 highest rated English language apps. All selected apps, independent of the language, were comparable in the following categories: *availability*, *functionality and design*, *ease of use*, and *information management and medical accuracy*. The English language apps scored significantly higher in the following categories: *potential for improving self-management*, *child-friendly*, *fun factor*, *learning factor*, and *range of function*. English language apps (mean total points 34.164, SD 1.09) performed significantly better than German language asthma apps (mean total points 22.91, SD 2.898; *P*=.003). The best rated English language app was *Kiss my asthma* (36/42 points), whereas the best rated German language app *Kata* achieved only 27.33 points.

Conclusions: The recommended English language apps are *Kiss my asthma*, *AsthmaXcel*, *AsthmaAustralia*, and *Ask Me, AsthMe!*, whereas the only recommended German language app is *Kata*. The use of apps plays an increasingly important role in patients' lives and in the medical field, making mHealth a staple in the future of asthma treatment plans. Although validated recommendations on rating mHealth apps have been published, it remains a challenging task for physicians and patients to choose a suitable app for each case, especially in non–English-speaking countries.

(JMIR Mhealth Uhealth 2021;9(5):e24907) doi:10.2196/24907

KEYWORDS

asthma; apps; mobile health; self-management; recommended apps; German; English; mobile phone

Introduction

Background

Asthma is the most common chronic disease in childhood, affecting 1 in 12 children. According to the World Health Organization, approximately 339 million [1] people worldwide are living with asthma. Asthma is not a harmless condition; according to the last World Health Organization survey in 2016, there were 417,918 deaths attributable to asthma [2]. Disease control is often difficult, especially in children, because of a poor understanding of the issue and misjudging the severity of the symptoms. According to the Global Initiative for Asthma, the treatment adjustment strategy includes education, skills training, and optimization of medications as pillars of personalized disease management. Mobile health (mHealth) apps have recently become a tool to better educate children about their illness and treatment management [3].

A systemic analysis by Farzandipour et al [4] proved the exceptional potential asthma apps have to improve the quality of life and control symptoms compared with conventional treatment methods. Routine care can benefit from functions such as medication intake reminders, symptoms tracking, transmission of peak flow measurements directly to the treating physician [5], and tailored education about the disease and its risks.

The use of apps has increased significantly in recent years and, accordingly, the use of medical apps. By 2020, the number of smartphone users worldwide is expected to reach 6.1 billion or 80% of the world's population [6]. Approximately 62% of all smartphone owners have used their smartphone to find information about their health [7]. According to a study conducted in Hong Kong, one of the most hi-tech cities, approximately 24% of smartphone and tablet users installed a mHealth app [8]. In 2015, the available mHealth apps in the relevant stores were more than 100,000, with almost 3 trillion downloaded mHealth apps [9]. In Europe, approximately 46% of children already had a smartphone in 2019 and 41% used it every day [10]. These figures draw attention to a rising market, making further research necessary. Although the classification of mHealth apps for asthma in English has already been the focus of several studies [11-13], research has neglected German-speaking countries. Multiple tools shortlist and evaluate apps that are potentially useful in the management of chronic diseases. Some also include a guide for rating or creating a standardized user dummy, thus minimizing interindividual effects of various testers [14-16].

In Germany, more than 11 million minors are aged <14 years [17]. The prevalence of asthma among children is approximately 10% [18], indicating that the market for potential users of asthma apps for children in Germany exceeds the 1 million mark.

Objectives

Therefore, this study aims to identify functional asthma apps in German and to compare them with English language apps. In a 3-step system, the *Google Play Store* and *Apple App Store* were systematically searched to preselect the most efficient apps using a criteria catalog, which was in part self-generated and in part adopted from existing validated catalogs. The content was further analyzed for correctness according to the current guidelines.

Methods

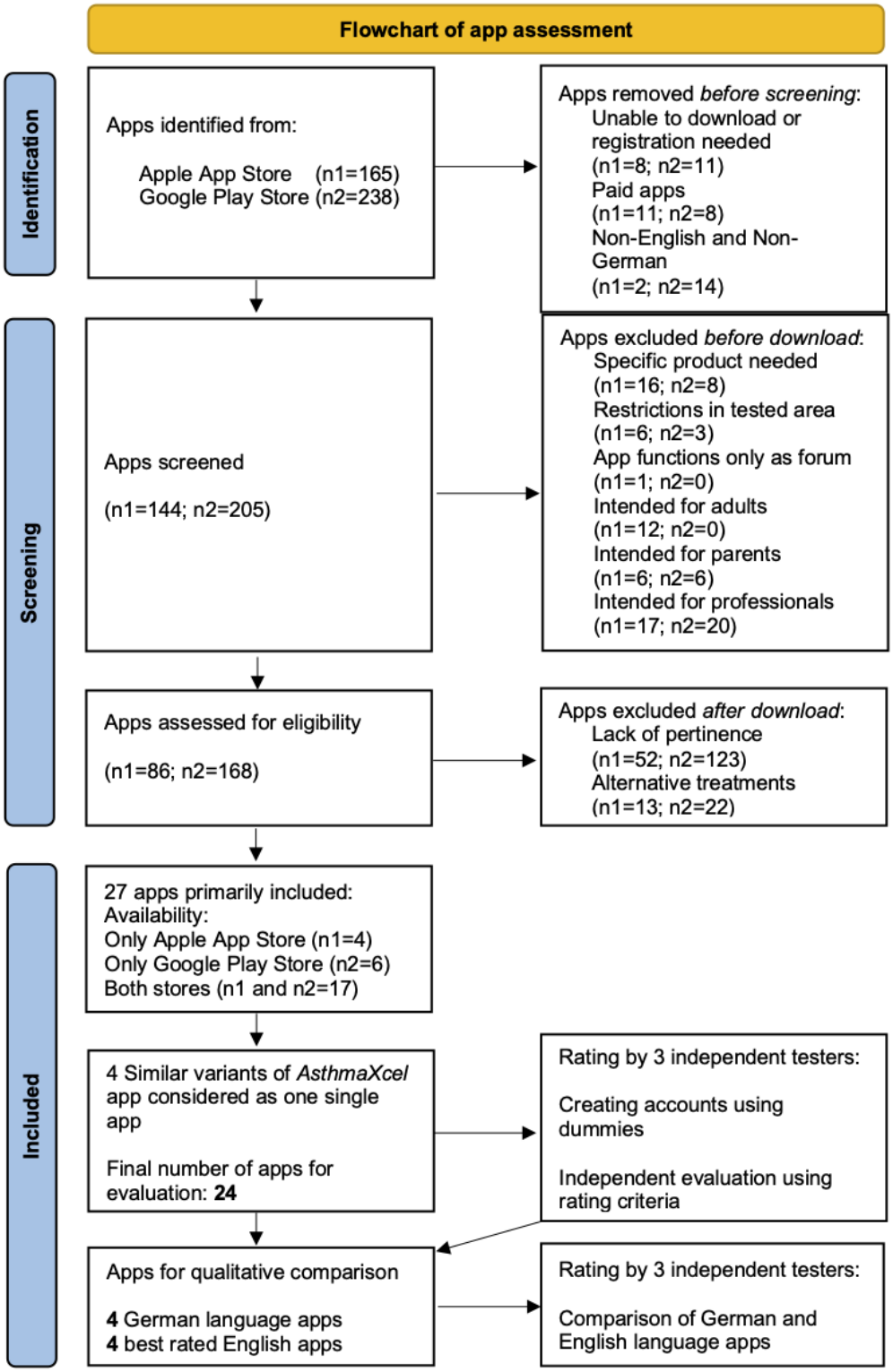
Quantitative Comparison of Apps

Two of the world's leading mobile app platforms, the Google Play Store and Apple App Store, were used for the search. The preselection process following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Figure 1), adapted from the study by Page et al [19], was performed by one of the authors according to the inclusion and exclusion criteria. In the case of questionable results, they were checked for validity and plausibility by the other authors to ensure an objective decision.

The term *asthma* was searched in both app stores, and the apps eligible for inclusion were downloaded on a smartphone with a Google Android operating system or an Apple iPad with an iOS operating system. Inclusion (available free of charge, German or English language apps, and suitable for children and adolescents) and exclusion criteria (specific product needed; restrictions in the tested area; app forum as the only function; intended for adults, parents, or medical professionals; alternative treatments; and lack of pertinence; Multimedia Appendix 1 [3,18,20]; Figure 1) were applied to select relevant apps and compare German and English free offers for children and adolescents. As asthma affects all social classes, we only included free apps. Other studies have already pointed out that low- and middle-income populations have limited access to mHealth apps [7]. For these groups, free apps were the only viable option.

As shown in Multimedia Appendix 1, these criteria ruled out all apps irrelevant for comparison and restricted the choice to those catering to children and adolescents, those available in German and English, those available free of charge, and available following medical guidelines and recommendations.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of app assessment. After applying the exclusion criteria, 27 apps were primarily included. Of these, 17 were available in both stores, 4 in the Apple App Store (depicted as n1), and 6 in the Google Play Store (depicted as n2). As 1 app (AsthmaXcel) had 4 different variants, it was considered as one single app, resulting in 24 apps for the final analysis.



Qualitative Assessment of German and English Language Apps

After the preselection process, all remaining apps that met the inclusion criteria were downloaded, and their functions were evaluated by 3 independent persons following a self-compiled criteria catalog. These 3 raters tested the apps independently and without any conflicts of interest.

For an objective analysis, testing was performed following the same procedure. A private email address and contact details were used for logging in, when required, to access the functions of the apps. For a realistic approach, any minor age was selected,

and, if necessary, realistic, arbitrary values about personal measurements and other data were selected.

In total, 2 test dummies (Multimedia Appendix 2) of different age groups (preschoolers and teenagers) were generated to test all available app features for each age group and gain a comprehensive insight into each app.

A total of 27 apps meeting the inclusion and exclusion criteria, suitable for use by children and adolescents, were selected for evaluation and representative comparison of their quality. The *AsthmaXcel* app with its 4 different, complementary, and overlapping variants (*AsthmaXcel*, *AsthmaXcel PRO*, *AsthmaXcel Adventures*, and *AsthmaXcelED*) was considered as one single app for the evaluation. Thus, the final number of evaluated apps was 24. The resulting app selection is presented in more detail in the *Results* section.

On authorization for this study, quality assessment was carried out by the 3 independent testers according to a criteria catalog consisting of 9 categories, some conceived for this purpose and some adopted from existing validated catalogs (Multimedia Appendix 3 [3,14-16,18,20]). The categories *availability*, *child-friendly*, *learning factor*, and *range of functions* were analyzed using self-compiled criteria. The categories *functionality and design*, *ease of use*, *potential for improving asthma self-management*, *fun factor and incentives*, and *information management and medical accuracy* were adopted from existing catalogs [14-16].

As there are no official criteria for evaluating apps, particularly for asthma apps for children, the authors designed some specifically intended to assess apps' suitability for children and integrated them with existing standards. The child-friendliness of an app was determined based on the provider's recommendations, the design (visual incentives for children to use the app), the range of functions (games and understandable information about asthma), engagement creation (through a reward or score system and entertaining features for children), usability (by the child alone or with parental support), and the general impression of the app. The test dummies were created to simulate a child using an app as accurately as possible (Multimedia Appendix 2).

The category *information management and medical accuracy* was analyzed using the *App Chronic DiseaseChecklist* by Anderson et al [16]. In addition, the quality of the app and its medical accuracy were checked by the authors using the Global Initiative for Asthma [3,20] and German Airway League [18] guidelines to ensure medical guidelines were met and if the recommendations (eg, instructional videos) were correct. This category was not designed to check every single guideline recommendation but rather to analyze the correctness of the provided information in general (eg, asthma medications, time point of adapting the medication, how to access medical care, and how to use an inhaler). These factors influenced the rating scores by consensus [14-16]. As 4 of the questions in this category addressed the points of information management, data protection, and app provider, this category was named *information management and medical accuracy*.

For the criteria catalog, we used a point system. Apps matching 8 categories were assigned a value between 1 and 5; those matching 1 category (availability) could score either 1 or 2 points, as illustrated in Multimedia Appendix 3. The validated criteria catalog point system differed from the one we applied to the *potential for improving asthma self-management* [14] and *fun factor and incentives* [15] categories (0-26 and 0-31 points, including half points for partial compliance) and to the *functionality and design* [16], *ease of use* [16], and *information management and medical accuracy* [16] categories (0-6 points, including half points for partial compliance). Multimedia Appendix 3 provides the ratings for each class. Classes indicate an app rating for a category with a different point system to align it with our 5-point system. The *App Chronic Disease Checklist version 1.0* according to Anderson et al [16] for the *functionality and design*, *ease of use*, and *information management and medical accuracy* [16] categories consists of 6 questions. Each response was assigned a value of 1 (full compliance), 0.5 (partial compliance), or 0 (no compliance), hence the maximum of 6 points. The points scored in the 6 questions were divided into 5 classes to ensure compatibility of the checklist protocol with our 5-point system. Half of the points were counted in the next higher class (Multimedia Appendix 3).

All 24 apps (4 versions of *AsthmaXcel* were considered as one and referred to as *AsthmaXcel*) that met the inclusion and exclusion criteria were evaluated, compared, and ranked based on the points achieved both cumulatively and in the single categories.

A total of 3 testers individually assessed the apps and then determined the mean value of points both cumulatively and in the single categories.

Therefore, apps were ranked according to the average of the total points of the 3 testers and the point system, with a maximum of 42 points. Statistical calculations on the same number of highest-ranking German and English language apps compared the quality of the apps. As only 4 of all eligible apps were in German, these were compared with the 4 highest ranked English language apps.

Statistical Analysis

The quantitative comparison between German and English language apps considered the corresponding offers on the Google Play Store and Apple App Store. The analysis was performed using Microsoft Excel and IBM SPSS Statistics 25.0. Descriptive statistics were used to calculate and compare the results. For normally distributed results, the mean was calculated with SD, and for nonnormally distributed results, the median was determined between the maximum and minimum values. Normal distribution was tested using the Shapiro-Wilk test and confirmed for total points and all categories except *availability*, for which other variables nonparametric tests were used. The significant difference in the results was calculated using the two-tailed *t* test or, in the case of nonparametric results, using the Mann-Whitney U test. Statistical significance was set at $P<.05$.

Results

Quantitative Comparison of German and English Language Asthma Apps for Children and Adolescents

In total, 403 apps were identified under the term *asthma* in both app stores (Figure 1), including 238 (59.1%) on the Google Play Store and 165 (40.9%) on the Apple App Store. Of 403 apps, 27 (6.7%) met the inclusion and exclusion criteria. Of these, 17 (17/27, 63%) were available in both stores, 4 (4/27, 15%) only on the Apple App Store, and 6 (6/27, 22%) only on the Google Play Store. As the *AsthmaXcel* app has 4 different, complementary, and overlapping variants (*AsthmaXcel*, *AsthmaXcel PRO*, *AsthmaXcel Adventures*, and *AsthmaXcelED*) separately available for download, they were considered as one single app for the evaluation. Therefore, the final number of apps downloaded for the analysis was 24.

Of these 24 apps, 20 (83%) were available in English but not in German, and 4 (17%) apps were available in both German and English. None of the apps were available only in German.

The most common exclusion criterion was the lack of pertinence to the topic, although it was listed under the term asthma. The reasons for exclusion are listed in Table 1 and Figure 1.

Table 1. Exclusion criteria for the asthma apps.

Exclusion criteria	Google Play Store (n=215), n (%)	Apple App Store (n=144), n (%)
Lack of pertinence	123 (57.2)	52 (36.1)
Alternative treatment methods	22 (10.2)	13 (9.0)
Language	14 (6.5)	2 (1.4)
Cost	8 (3.7)	11 (7.6)
Intended for medical staff	20 (9.3)	17 (11.8)
Intended for parents	6 (2.8)	6 (4.2)
Impossible download, registration, or use	11 (5.1)	8 (5.5)
Medical device or specific product required	8 (3.7)	16 (11.1)
Asthma forum	0 (0)	1 (0.7)
Limited functions in some areas	3 (1.4)	6 (4.1)
Intended for adults	0 (0)	12 (8.3)

Quality Assessment

All 24 apps that met the inclusion and exclusion criteria were rated by the 3 app testers. Average single category point values and total points were calculated (Table 2) and ranked based on the average total points. The calculation of the average of the points awarded by all 3 testers sometimes resulted in decimal values that were not rounded for better differentiation in the apps' performance. Multimedia Appendix 4 shows the evaluation of each tester.

Table 2. Average of the 3 testers' evaluation (individual categories and overall points of English and German language apps).

App	Categories[a]										
	Language	Availability	Functionality and design	Ease of use	Potential for improving asthma self-management	Child-friendly	Fun factor and incentives	Learning factor	Information management and medical accuracy	Range of function	Total
Maximum points	__[b]	2.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	42.0
KmAsthma	English	2.0	5.0	5.0	4.0	3.3	3.0	4.7	4.3	4.7	36.0
AsthmaXcel	English	2.0	4.0	4.3	3.3	4.7	3.3	5.0	4.0	4.0	34.6
Asthma Australia	English	2.0	4.7	3.7	3.0	5.0	2.3	4.7	3.7	4.3	33.4
Ask Me, AsthMe!	English	2.0	5.0	4.7	4.0	4.3	2.0	4.0	3.7	3.7	33.4
AsthmaMD	English	2.0	4.7	5.0	3.3	3.0	2.0	4.3	4.7	4.0	33.0
Elfy	English	2.0	4.7	4.7	2.3	3.3	1.0	4.0	4.0	3.3	29.3
Kata	German and English	2.0	5.0	3.7	3.0	2.0	1.0	2.3	5.0	3.3	27.3
Wizdypets	English	2.0	3.7	3.3	2.3	5.0	3.0	3.0	2.3	2.3	26.9
Asthmadodge	English	2.0	3.3	2.7	1.7	5.0	2.3	3.3	2.7	2.7	25.7
Asthma Eclub	English	1.0[c]	3.3	3.0	2.0	4.3	1.0	4.7	3.7	2.3	25.3
SaniQ	German and English	2.0	4.0	4.3	1.7	2.3	1.3	1.7	3.3	3.0	23.6
Asthma Tracker	German and English	2.0	3.3	3.7	1.7	2.0	1.0	1.0	3.3	2.7	20.7
Breathcount	English	1.0[d]	4.3	3.7	1.7	2.3	1.3	1.7	2.7	1.7	20.4
Allergymonitor	German and English	2.0	3.0	3.3	1.0	2.7	1.0	2.3	2.7	2.0	20.0
Rightbreath	English	1.0[c]	3.0	3.0	1.7	2.0	1.0	3.0	2.7	2.3	19.7
InhalerCounter	English	1.0[c]	3.7	3.7	1.7	1.7	1.0	1.7	2.3	2.7	19.5
AsthmaActionhero	English	1.0[c]	3.7	3.0	1.0	4.0	1.0	1.7	2.3	1.7	19.4
ASTHMA	English	1.0[c]	3.7	3.3	1.7	2.3	1.0	2.0	1.7	2.0	18.7
Asthma:Management	English	1.0[d]	1.7	3.3	1.7	1.7	1.0	3.0	2.7	1.7	17.8
mypeakflow	English	1.0[d]	3.3	2.7	1.3	1.7	1.0	1.7	3.0	1.7	17.4
Inhaler	English	1.0[d]	1.7	2.0	1.0	2.7	1.0	2.3	3.0	1.7	16.4
Peak Flow	English	1.0[d]	3.3	2.0	1.7	1.7	1.0	1.3	2.7	1.7	16.4
Asthma	English	1.0[d]	2.3	2.0	1.3	1.7	1.0	2.7	2.3	2.0	16.3
Inhaler diary	English	1.0[d]	2.7	2.3	1.0	1.3	1.0	1.3	2.0	1.3	14.0

[a]Average points of the 3 testers presented with one decimal.

[b]Cells do not add points to the scoring system but reflect the available language (English or German) and the score provided at the end after applying the total points.

[c]Apple.

[d]Android.

Qualitative Comparison of German and English Language Asthma Apps

The best rated apps (Textbox 1) in English and German were selected and used for the statistical comparison of their quality. As only 4 of all eligible apps were in German, these were compared with the 4 highest ranked English language apps.

Quality was evaluated and compared according to the points assigned cumulatively and in single categories of the criteria catalog (Table 2).

Textbox 1. Four best rated apps in English and German.

Apps in German

- SaniQ Asthma
- Asthma Tracker
- Kata
- AllergyMonitor

Apps in English

- Kiss my asthma (KmAsthma)
- AsthmaXcel
- Asthma Australia
- Ask Me, AsthMe!

Categorial Analysis of Asthma Apps

Availability

As all 4 English language apps and 4 German language apps were available on the Google Play Store and the Apple App Store, they were awarded a maximum score of 2 points (median 2, minimum 2, and maximum 2) without showing any differences.

Functionality and Design

For the evaluation of this category, the *functionality* subitem in the *App Chronic Disease Checklist version 1.0* according to Anderson et al [16] was used. The English language apps *KmAsthma* and *Ask Me, AsthMe!* and the German language app *Kata* scored the highest (5 points). Points were deducted for other apps because of the lack of the possibility of sending data such as peak flow values and the lack of warnings for out-of-range values. The design and performance power were very satisfactory in all 8 apps; hence, all apps received at least three points in this category. The average score of all 8 apps was 4.25 (SD 0.7), the average score of apps in English was 4.7 (SD 0.41) and the average score of apps in German was 3.8 (SD 0.76). The difference between the German and English language apps was not significant (P=.16).

Ease of Use

The *App Chronic Disease Checklist version 1.0* according to Anderson et al was also used for the evaluation of this category [16]. *Km Asthma* was the only app that achieved 5 points in the *ease of use* subitem [16], as it is the only one that allows users to perform all self-management tasks easily, creates a user profile with a log-in option, and offers a reminder function. Points were deducted from all other apps for not offering these features. Nevertheless, all apps in this category performed very well, with a minimum score of 3.33, as all apps can be used offline and are intuitive. The average score for all 8 apps was 4.08 (SD 0.54), with the English language apps performing slightly, but not significantly, better (P=.11). The average of the English language apps was 4.4 (SD 0.49), and of the German language apps was 3.74 (SD 0.34).

Potential for Improving Asthma Self-management

Evaluation for this category was carried out with the help of the "Exemplary rating criteria for behavior change techniques in mHealth asthma apps" according to Abraham and Michie [14].

On the basis of the points achieved, 5 classes were conceived to align the rating in this category with the alternative 1-5-point system. None of the apps were awarded 5 points in this category. The apps *Km Asthma* and *Ask Me, AsthMe!* achieved 4 points in this category and thus the highest score among all apps. The German language app *AllergyMonitor* only achieved a minimum score of 1 point in this category. English language apps performed significantly better (P=.02), with an average score of 3.6 (SD 0.44), whereas German language apps' average score was 1.8 (SD 0.72). The mean score of all 8 apps was 2.7 (SD 1.05).

Child-Friendly Factor

The results diverged widely in this category. Only the English language app *Asthma Australia* achieved the highest score, closely followed by *AsthmaXcel* with 4.67. A distinctive element of these 2 apps was the age-appropriate educational videos on asthma and its treatment. German language apps received less than 3 points in this category. Among these, *AllergyMonitor* was the best rated in this category, with 2.67 points. The principal reasons for points deduction were the lacking content and functions suitable for children (*Kata*, *AllergyMonitor*, *Asthma Tracker*, and *SaniQAsthma*) along with a poorly captivating design for the younger ones (*Kata* and *SaniQAsthma*). The average score was 3.28 (SD 1.15), with a significant difference (P=.007 between English language apps (average score 4.3, SD 0.63) and German language apps (average score 2.2, SD 0.3).

Fun Factor and Incentive to Use the App

This category was assessed using the "Exemplary rating criteria for gamification components in mHealth asthma apps" according to Thiebes et al [15]. On the basis of the points achieved, 5 classes were conceived to align the rating in this category with the alternative 1- to 5-point system. The app *AsthmaXcel* scored 3.33 points in this category, the highest among all apps.

The average of all 8 apps was only 1.87 (SD 0.88). Although 3 of the 4 German language apps received only 1 point, with a mean value of 1.075 (SD 0.13), the mean value of the English language apps was 2.67 (SD 0.53), indicating a significant difference between the 2 language groups (P=.02).

Learning Factor

Only *AsthmaXcel* achieved 5 points because videos about asthma and its treatment and games with questions about the topics contribute to the app's learning factor. In this category, points were deducted because of inadequate information (*Km Asthma*, *Ask Me*, *AsthMe!*, and *AllergyMonitor*), no child-friendly delivery of information (*Kata*) or lacking content (*Asthma Tracker* and *SaniQAsthma*) about asthma and its treatment. The mean score of the 8 apps in this category was 3.2 (SD 1.45). The mean value of apps in German language was 1.83 (SD 0.55), whereas the mean value of apps in English language was 4.59 (SD 0.36). With P<.001, there was a significant difference in this category between the 2 language groups.

Information Management and Medical Accuracy

The accuracy of medical content is undoubtedly crucial [3,18,20]. This category was the only one with a German app (*Kata*) outperforming the English apps for following the *German Respiratory Society* website medical guidelines [18] and providing references. Points were deducted because of a lack of references (*Asthma Australia*, *Km Asthma*, and *AsthmaXcel*). An additional point was deducted for *Ask Me*, *AsthMe!*, *AllergyMonitor*, *Asthma Tracker*, and *SaniQAsthma* because of poor content and missing references. The average score of all 8 apps was 3.7 points (SD 0.63). The mean value of English language apps was 3.84 (SD 0.165), whereas the mean value of German language apps was 3.58 (SD 0.86). The difference in points achieved in this category by English and German language apps was not significant (P=.65).

Range of Functions

As none of the apps incorporated all the beneficial functions (general asthma information, games, diary, medication reminder, and pollen calendar), none of them achieved the highest score. The gaming app *AsthmaXcel* and the information app *Asthma Australia* lack diary and medication reminder functions, whereas diary apps (*Km Asthma*, *Ask Me, AsthMe!*, *Kata*, *AllergyMonitor*, *Asthma Tracker*, and *SaniQAsthma*) do not provide a game function.

The broadest array of functions was offered by the apps *Km Asthma* (diary and medication reminder functions, asthma action plan, and information), *AsthmaXcel* (games, information, and video quiz), and *Asthma Australia* (information, videos, quizzes, asthma control tests, and asthma action plans). These apps scored more than 4 out of 5 points. The best rated German app *Kata* (diary, medication reminder, asthma control test, inhalation instructions, pollen calendar, and general asthma information) achieved 3.33 points. Among the English language apps, most points were deducted in this category for the *Ask Me, AsthMe!* app (diary, asthma action plan, asthma control test score, inhalation videos, and inadequate general information). The German language apps *Asthma Tracker* (diary, medication reminder, and asthma control test) and *SaniQAsthma* (diary and medication reminder) scored 3 and 2.67 points, respectively, whereas *AllergyMonitor* incorporated the diary function only, thus totaling 2 points.

The mean value of the 8 apps in this category was 3.46 (SD 0.82). The mean value of the English language apps was 4.16 (SD 0.37), and the mean value of the German language apps was 2.75 (SD 0.49). The difference in points achieved between the 2 language groups was significant (P=.007).

Total Points and Overall Ranking

The evaluation of the quality of the apps using the criteria catalog determined that the English language asthma apps performed significantly better than the German language apps (Table 3).

Table 3. Total score of English and German language apps.

Rank	App	Language	Points, mean (SD)
1	*KmAsthma*	English	36.00 (0.82)
2	*AsthmaXcel*	English	34.00 (1.41)
3	*Asthma Australia*	English	33.33 (1.25)
4	*Ask me, Asthme!*	English	33.33 (1.25)
5	*Kata*	German	27.33 (0.94)
6	*SaniQ*	German	23.67 (1.25)
7	*Asthma tracker*	German	20.67 (4.03)
8	*Allergymonitor*	German	20.00 (2.94)

Km Asthma was the best rated app with 36 out of 42 points, followed by *AsthmaXcel* with 34 out of 42 points. *Km Asthma's* unique characteristics are personalized asthma-related goals, such as medication reminders; the possibility to learn about trigger factors; and the animated, funny characters, making this app particularly suitable for children. The app also offers an asthma calendar, information about the disease, and a user-specific asthma action plan, facilitating asthma self-management. A noteworthy aspect is that *Km Asthma* and *Asthma Australia* have the same developers. *AsthmaXcel* distincts itself for its updated version incorporating both informative content and gaming components. There are 4

AsthmaXcel versions (*AsthmaXcel*, *AsthmaXcel PRO*, *AsthmaXcel Adventures*, and *AsthmaXcelED*). *AsthmaXcel*, *AsthmaXcel PRO*, and *AsthmaXcelED* are information apps all containing the same animated videos about asthma. *AsthmaXcel PRO* also offers virtual coins as a reward for correctly answered questions about the videos and a leaderboard ranking all users. *AsthmaXcel Adventures* is meant as a supplementary app with games designed to answer questions about other *AsthmaXcel* apps videos to get life points. This combination achieves an exceptionally high learning effect, with a focus on child-friendly education. The best rated German language app (*Kata*) achieved 27.33 out of 42 points. The app is user-friendly and has a wide range of functions (diary, medication reminder, asthma control test, inhalation instructions, pollen calendar, and general asthma information). However, the app is only partially child-friendly, and its results in the *fun factor and incentives* category were not particularly convincing.

The mean value of all the 8 apps was 28.54 points (SD 6.03). The average of the 4 English language apps was 34.165 points (SD 1.09). The average of the 4 German language apps was 22.91 points (SD 2.898). All English language apps ranked above the overall mean, whereas German language apps were below that value. The difference in the total number of points in the criteria catalog between German and English language apps was highly significant (P=.01).

Discussion

Principal Findings

This study aims to identify the quality and quantity of mHealth apps for English- and German-speaking children with bronchial asthma. Our analysis and evaluation following a criteria catalog that was in part self-compiled and in part applying analysis scores for mHealth apps [14-16] confirmed the qualitative and quantitative inequality between German and English language asthma apps. This is not surprising, as English-speaking countries are generally ahead as far as apps development is concerned. Although the variety of English language asthma apps is extensive, only a fraction of apps are available in German. The 4 apps in German language performed significantly worse than the 4 apps in English language, with an average score of 22.91 (SD 2.898) for the former versus 34.165 (SD 1.09) for the latter. The best rated English language app (*Km Asthma* 36 points) received a substantially higher score than the best rated German language app (*Kata* 27.33 points). A striking contrast was evident in 7 of the 9 categories. The German language apps only performed equally well in the *availability* category, and the *Kata* app was the best performing in the *information management and medical accuracy* category.

This study is the first to compare German and English language asthma apps for children and adolescents, and generally, little information exists about asthma apps dedicated to these age groups or, more specifically, about apps developed in German. The assortment of asthma apps on the Google Play Store and Apple App Store is large, but only a few are child-friendly, and providing comprehensive medical recommendations remains a difficult task, as none of the available options integrate all the necessary functions. As other studies also underlined, asthma self-management requires a combination of at least two apps to access all features [12,21]. Functions range from educative games to asthma diary, transmission of symptoms and peak flow values directly to the attending physician, and pollen calendars. This variety is essential, and, for children, in particular, apps that playfully educate children about asthma are recommended.

Children with high-risk asthma seem to be inclined to use asthma apps [22]. However, the search for an asthma app that pediatric patients can use on a daily basis for disease self-management in coordination with the treatment plan represents a challenge for children, adolescents, and their parents, which is why physicians rarely integrate their use. An American study refuted the concerns related to the transmission of personal health data via apps between patients and physicians [23]. Data protection of personal data is a concern for many patients and parents. Data protection regulations in English-speaking countries, such as the United States, and the European Union are quite diverse, with the European Union having more restrictive regulations. These restrictions might complicate mHealth development. As medical professionals often fail to provide exhaustive information about this tool, the choice falls on the patients or their caregivers.

However, mHealth apps are not suitable for all patient groups. Not all families might have the financial means of purchasing a smartphone for their children. Social and language barriers may hinder the use of mHealth apps. These patient groups will benefit from continuing to use conventional, manually recorded asthma diaries.

The approach to using the apps typically differs depending on age group. Primary school children will prefer playful apps that can significantly nurture their enthusiasm and motivation to learn. For instance, *AsthmaXcel* creates incentives through fun games. Currently, there are no German language apps that integrate games. The *AsthmaXcel* app was already the subject of a study about its educational aspect, but there are no confirmed results [24]. However, since the launch of its updated, more informative versions, *AsthmaXcel PRO* and *AsthmaXcel Adventures*, further studies should be conducted. *Asthma Australia* teaches children the basics of their illness using engaging and clear animations by their peers. For primary school–aged children who can already read, the *Km Asthma* and *Ask Me, AsthMe!* apps are a valid alternative as the only 2 available *diary apps*, both in German and English. Tracking asthma-related statistics with the help of a *diary app* can significantly improve disease management [25]. Regrettably, no German language app is suitable for younger primary school pupils. For older children and adolescents, the *AsthmaXcel (Adventures)* app can be recommended, as the game with questions is a useful tool for this age group. However, no gaming app for this specific age group is available in German language. As for the English language apps' diary function, *Km Asthma* and *Ask Me, AsthMe!* are the best choice for primary school children, whereas *Km Asthma* is the best option for teenagers. *Asthma Australia* is a valid recommendation also for older children and adolescents, as the informative videos section is developed considering those specific cohorts. The *Kata* app is best suited for German speakers because it connects to the

German Respiratory Society website through a direct in-app hyperlink, which is an excellent source of information on asthma. All 4 German language apps (*Kata*, *SaniQ*, *Asthma Tracker*, and *Allergy Monitor*) offer teenagers the option of keeping an asthma diary. These are not suitable for younger children, as they hardly offer child-friendly content. According to the developer, the *SaniQ* app is intended for users aged ≥16 years, but the App Store description recommends it from 4 years of age, therefore meeting the age-related inclusion parameter for this study.

Studies conducted on numerous diseases have already confirmed the potential of apps to support self-management and be instrumental in the treatment plan [26-28]. An increasing number of studies are focusing on the use of asthma apps [4,29-32]. Controversial research data on asthma apps' influence on symptom improvement and exacerbation frequency reduction are also available. Symptoms appeared to improve using asthma apps [33-35], with no difference compared with placebo [33,36]. However, almost all studies investigating asthma apps included mainly adult study participants [4,10] and few adolescents [30,31,33,37-39]. A separate study concentrating on the benefits of a particular asthma app reported an improved management of the disease through the app, but with no significant difference in the control group [39]. However, the app was used by the child's legal guardians.

Considering the growing mHealth apps market in German-speaking countries, a German version of the *Mobile App Rating Scale* [40], the *MARS-G* [41], was released in March 2020, after the final planning of our study results. Although it was not integrated into this study, *MARS-G* can be a tool for future studies on German language asthma apps.

Some studies have also evaluated the quality of asthma apps and issued recommendations for the use of suitable asthma apps [11,13,42,43]. However, data to compare the quality of asthma apps in German and English and their suitability for use by children remain lacking.

Limitations

Although app evaluation followed objective criteria, the ratings in each category were determined by 3 persons, thus the likelihood of a biased perspective. Moreover, apps were used over a limited period, and potential updates or apps issued after the completion of this study were not included. Four apps for each language were selected as the most suitable for quality comparison; however, as the array of apps for German speakers was limited, apps in German in addition to English were included in this group.

As asthma is a disease that affects all social classes worldwide, only free-of-charge apps were included in the evaluation. This choice represents a limitation because a free app typically does not have a full range of functions and a corresponding technical implementation.

As not all apps indicate an age limit, the app provider recommendation was used to select the apps. As adolescents are considered to be children up to the age of 17 years, all apps that are not explicitly labeled as intended for adults were included. Another limiting factor is that teenagers' preferences tend to be closer to those of adults' preferences. Hence, the apps were rated based on their suitability for toddlers and school-aged children.

As not every app offers an asthma diary with the entry of peak flow values, the diary feature was not defined as a separate category, and warnings for out-of-range values were not tested. This aspect was included as a question in the *functionality and design* category. Separate testing of asthma diary apps should be carried out to address this limitation. As neither the Apple App Store nor the Google Play Store provides German-only apps that meet both inclusion and exclusion criteria for this study, the apps supporting other languages including German were considered eligible for comparison.

Every German language app in this study is also available in English.

Conclusions

The use of apps plays an increasingly important role in patients' lives and in the medical field, making mHealth a staple in the future of asthma treatment plans. Although validated recommendations on rating mHealth apps have been published, it remains a challenging task for physicians and patients to choose a suitable app for each case, especially in non–English-speaking countries. Hence, further studies are required on this topic. In addition, developers should address the necessary features' improvements to allow a more efficient use of this tool in the future.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion and exclusion criteria or identification of suitable asthma apps.
[DOCX File , 13 KB - mhealth_v9i5e24907_app1.docx]

Multimedia Appendix 2
Test dummies for preschoolers and teenagers.
[DOCX File , 13 KB - mhealth_v9i5e24907_app2.docx]

Multimedia Appendix 3

Criteria catalog for each category and its description and rating.
[DOCX File , 18 KB - mhealth_v9i5e24907_app3.docx]

Multimedia Appendix 4
Single test results of each rater.
[DOCX File , 20 KB - mhealth_v9i5e24907_app4.docx]

References

1. Chronic respiratory diseases: asthma. World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/asthma [accessed 2021-05-04]
2. Asthma. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/asthma [accessed 2021-05-04]
3. Global strategy for asthma management and prevention. Global Initiative for Asthma. 2019. URL: https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf [accessed 2021-05-04]
4. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient self-management of asthma using mobile health applications: a systematic review of the functionalities and effects. Appl Clin Inform 2017 Dec;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
5. Wu AC, Carpenter JF, Himes BE. Mobile health applications for asthma. J Allergy Clin Immunol Pract 2015;3(3):446-448. [doi: 10.1016/j.jaip.2014.12.011] [Medline: 25725939]
6. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):106-115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
7. Stephan LS, Dytz AE, Guimaraes RB, Ley AG, Mathias RG, Assis MV, et al. Processes and recommendations for creating mHealth apps for low-income populations. JMIR Mhealth Uhealth 2017 Apr 03;5(4):e41 [FREE Full text] [doi: 10.2196/mhealth.6510] [Medline: 28373155]
8. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SS, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. JMIR Mhealth Uhealth 2017 Jun 05;5(6):e77 [FREE Full text] [doi: 10.2196/mhealth.7628] [Medline: 28583905]
9. Fischer F, Kleen S. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review. J Med Internet Res 2021 Jan 22;23(1):e17691 [FREE Full text] [doi: 10.2196/17691] [Medline: 33480850]
10. Licari A, Ferrante G, Marseglia GL, Giovanni C, Grutta SL. What is the impact of innovative electronic health interventions in improving treatment adherence in asthma? The pediatric perspective. J Allergy Clin Immunol Pract 2019 Nov;7(8):2574-2579. [doi: 10.1016/j.jaip.2019.08.008] [Medline: 31425835]
11. Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019 Nov;7(8):2583-2591 [FREE Full text] [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
12. Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. BMC Med 2012;10:144 [FREE Full text] [doi: 10.1186/1741-7015-10-144] [Medline: 23171675]
13. Tan R, Cvetkovski B, Kritikos V, O'Hehir RE, Lourenço O, Bousquet J, et al. Identifying an effective mobile health application for the self-management of allergic rhinitis and asthma in Australia. J Asthma 2020 Oct 24;57(10):1128-1139. [doi: 10.1080/02770903.2019.1640728] [Medline: 31274044]
14. 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]
15. Thiebes S, Lins S, Basten D. Gamifying information systems - a synthesis of gamification mechanics and dynamics. AIS eLibrary. 2014. URL: http://aisel.aisnet.org/ecis2014/proceedings/track01/4/ [accessed 2021-05-04]
16. Anderson K, Burford O, Emmerton L. App chronic disease checklist: protocol to evaluate mobile apps for chronic disease self-management. JMIR Res Protoc 2016 Nov 04;5(4):e204 [FREE Full text] [doi: 10.2196/resprot.6194] [Medline: 27815233]
17. Rudnicka J. Number of children under the age of 14 in Germany 2015-2019. Statista. 2020. URL: https://de.statista.com/statistik/daten/studie/1253/umfrage/anzahl-der-kinder-bis-14-jahre-in-deutschland-seit-dem-jahr-1950/ [accessed 2021-05-04]
18. Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V., Deutsche Gesellschaft für Rehabilitationswissenschaften e.V., und Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e.V., Buhl R, Bals R, Baur X, et al. [Guideline for the diagnosis and treatment of asthma - guideline of the German Respiratory Society and the German Atemwegsliga in cooperation with the Paediatric Respiratory Society and the Austrian Society of Pneumology]. Pneumologie 2017 Dec 07;71(12):849-919 [FREE Full text] [doi: 10.1055/s-0043-119504] [Medline: 29216678]
19. Page M, McKenzie J, Bossyut P. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br Med J 2021:372. [doi: 10.1136/bmj.n71]

20. Global strategy for asthma management and prevention. Global Initiative for Asthma. 2018. URL: https://ginasthma.org/wp-content/uploads/2018/04/wms-GINA-2018-report-V1.3-002.pdf [accessed 2021-05-05]
21. Hui CY, Walton R, McKinstry B, Pinnock H. Time to change the paradigm? A mixed method study of the preferred and potential features of an asthma self-management app. Health Informatics J 2020 Jun 13;26(2):862-879 [FREE Full text] [doi: 10.1177/1460458219853381] [Medline: 31195877]
22. Nichols M, Miller S, Treiber F, Ruggiero K, Dawley E, Teufel Ii R. Patient and parent perspectives on improving pediatric asthma self-management through a mobile health intervention: pilot study. JMIR Form Res 2020 Jul 03;4(7):e15295 [FREE Full text] [doi: 10.2196/15295] [Medline: 32442127]
23. Savage M, Savage L. Doctors routinely share health data electronically under HIPAA, and sharing with patients and patients' third-party health apps is consistent: interoperability and privacy analysis. J Med Internet Res 2020 Sep 02;22(9):e19818 [FREE Full text] [doi: 10.2196/19818] [Medline: 32876582]
24. Hsia B, Mowrey W, Keskin T, Wu S, Aita R, Kwak L, et al. Developing and pilot testing ASTHMAXcel, a mobile app for adults with asthma. J Asthma 2020 Feb 19:1-14. [doi: 10.1080/02770903.2020.1728770] [Medline: 32046564]
25. Venkataramanan R, Thirunarayan K, Jaimini U, Kadariya D, Yip HY, Kalra M, et al. Determination of personalized asthma triggers from multimodal sensing and a mobile app: observational study. JMIR Pediatr Parent 2019 Jun 27;2(1):e14300 [FREE Full text] [doi: 10.2196/14300] [Medline: 31518318]
26. Fruhwirth V, Enzinger C, Weiss E, Schwerdtfeger A, Gattringer T, Pinter D. [Use of smartphone apps in secondary stroke prevention]. Wien Med Wochenschr 2020 Feb 18;170(1-2):41-54. [doi: 10.1007/s10354-019-00707-3] [Medline: 31535230]
27. Liu S, Feng W, Chhatbar PY, Liu Y, Ji X, Ovbiagele B. Mobile health as a viable strategy to enhance stroke risk factor control: a systematic review and meta-analysis. J Neurol Sci 2017 Jul 15;378:140-145 [FREE Full text] [doi: 10.1016/j.jns.2017.04.050] [Medline: 28566151]
28. Klee P, Bussien C, Castellsague M, Combescure C, Dirlewanger M, Girardin C, et al. An intervention by a patient-designed Do-It-Yourself mobile device app reduces HbA1C in children and adolescents with type 1 diabetes: a randomized double-crossover study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]
29. Huang X, Matricardi PM. Allergy and asthma care in the mobile phone era. Clin Rev Allergy Immunol 2019 Apr 21;56(2):161-173. [doi: 10.1007/s12016-016-8542-y] [Medline: 27209270]
30. Jácome C, Almeida R, Pereira A, Araújo L, Correia M, Pereira M, INSPIRERS group. Asthma app use and interest among patients with asthma: a multicenter study. J Investig Allergol Clin Immunol 2020 Apr 23;30(2):137-140 [FREE Full text] [doi: 10.18176/jiaci.0456] [Medline: 32327403]
31. Wu AC. The promise of improving asthma control using mobile health. J Allergy Clin Immunol Pract 2016;4(4):738-739. [doi: 10.1016/j.jaip.2016.04.003] [Medline: 27393782]
32. Alquran A, Lambert K, Farouque A, Holland A, Davies J, Lampugnani E, et al. Smartphone applications for encouraging asthma self-management in adolescents: a systematic review. Int J Environ Res Public Health 2018 Oct 29;15(11):2403 [FREE Full text] [doi: 10.3390/ijerph15112403] [Medline: 30380692]
33. Liu W, Huang C, Wang C, Lee K, Lin S, Kuo H. A mobile telephone-based interactive self-care system improves asthma control. Eur Respir J 2011 Feb;37(2):310-317 [FREE Full text] [doi: 10.1183/09031936.00000810] [Medline: 20562122]
34. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632. [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
35. Cingi C, Yorgancioglu A, Cingi CC, Oguzulgen K, Muluk NB, Ulusoy S, et al. The "physician on call patient engagement trial" (POPET): measuring the impact of a mobile patient engagement application on health outcomes and quality of life in allergic rhinitis and asthma patients. Int Forum Allergy Rhinol 2015 Jun;5(6):487-497. [doi: 10.1002/alr.21468] [Medline: 25856270]
36. Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. Br Med J 2012;344:e1756 [FREE Full text] [Medline: 22446569]
37. Davis S, Peters D, Calvo R, Sawyer SM, Foster JM, Smith LD. A consumer designed smartphone app for young people with asthma: pilot of engagement and acceptability. J Asthma 2021 Feb;58(2):253-261. [doi: 10.1080/02770903.2019.1680997] [Medline: 31691618]
38. Fedele DA, McConville A, Graham TJ, McQuaid EL, Janicke DM, Turner EM, et al. Applying Interactive Mobile health to Asthma Care in Teens (AIM2ACT): development and design of a randomized controlled trial. Contemp Clin Trials 2018 Jan;64:230-237. [doi: 10.1016/j.cct.2017.09.007] [Medline: 28986245]
39. Real FJ, Beck AF, DeBlasio D, Zackoff M, Henize A, Xu Y, et al. Dose matters: a smartphone application to improve asthma control among patients at an urban pediatric primary care clinic. Games Health J 2019 Jun 03:357-365. [doi: 10.1089/g4h.2019.0011] [Medline: 31157983]
40. Stoyanov S, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

41. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
42. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113. [doi: 10.2196/mhealth.7177] [Medline: 28768606]
43. Kagen S, Garland A. Asthma and allergy mobile apps in 2018. Curr Allergy Asthma Rep 2019 Feb 02;19(1):6 [FREE Full text] [doi: 10.1007/s11882-019-0840-z] [Medline: 30712150]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 09.10.20; peer-reviewed by M Aksoy, K Anderson, V Strotbaum; comments to author 04.12.20; revised version received 18.02.21; accepted 16.04.21; published 19.05.21.

Please cite as:
Franzmair J, Diesner-Treiber SC, Voitl JJM, Voitl P
Effective German and English Language mHealth Apps for Self-management of Bronchial Asthma in Children and Adolescents: Comparison Study
JMIR Mhealth Uhealth 2021;9(5):e24907
URL: https://mhealth.jmir.org/2021/5/e24907
doi:10.2196/24907
PMID:34009134

Original Paper

Self-Management Apps for People With Epilepsy: Systematic Analysis

Mohsen Zaied Alzamanan[1], MS; Kheng-Seang Lim[2], PhD; Maizatul Akmar Ismail[3], PhD; Norjihan Abdul Ghani[3], PhD

[1]Department of Information Systems, University of Malaya, Kuala Lumpur, Malaysia

[2]Division of Neurology, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[3]Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Maizatul Akmar Ismail, PhD
Department of Information Systems
Faculty of Computer Science and Information Technology
University of Malaya
Jln Profesor Diraja Ungku Aziz
Kuala Lumpur, 50603
Malaysia
Phone: 60 133476676
Email: maizatul@um.edu.my

Abstract

Background: Patients with epilepsy (PWEs) are motivated to manage and cope with their disorder themselves (ie, self-management [SM] is encouraged). Mobile health (mHealth) apps have multiple features that have a huge potential to improve SM of individuals with chronic disorders such as epilepsy.

Objective: This study aimed to review all freely available apps related to the SM of PWEs and to determine the SM domains covered in these apps.

Methods: We performed a search of apps on Google Play and App Store using the keywords “epilepsy” or “seizures” from May to August 2018. Apps were included if they were free and in English language. We excluded apps with installation-related issues and not related to epilepsy self-management (eSM).

Results: A total of 22 eSM apps were identified in our search: 6 of these run only on iOS, 7 only on Android, and 9 run on both operating systems. Of the 11 domains of SM, seizure tracking and seizure response features were covered by most apps (n=22 and n=19, respectively), followed by treatment management (n=17) and medication adherence (n=15). Three apps (Epilepsy Journal, Epilepsy Tool Kit, and EpiDiary) were installed more than 10,000 times, with features focused specifically on a few domains (treatment management, medication adherence, health care communication, and seizure tracking). Two apps (Young Epilepsy and E-Epilepsy Inclusion) covered more than 6 SM domains but both had lower installation rates (5000+ and 100+, respectively).

Conclusions: Both Android and iOS mHealth apps are available to improve SM in epilepsy, but the installation rate of most apps remains low. The SM features of these apps were different from one another, making it difficult to recommend a single app that completely fulfills the needs of PWEs. The common features of the apps evaluated included seizure tracking and seizure response. To improve the efficacy and availability of these apps, we propose the following: (1) involve the stakeholders, such as physicians, pharmacists, and PWEs, during the development of mHealth apps; (2) assess the efficacy and acceptance of the apps objectively by performing a usability analysis; and (3) promote the apps so that they benefit more PWEs.

(JMIR Mhealth Uhealth 2021;9(5):e22489) doi:10.2196/22489

KEYWORDS

mobile health; epilepsy; self-management; smartphone

Introduction

Epilepsy affects approximately 50 million people of all ages worldwide, irrespective of race and economic background [1,2]. It is estimated that nearly 200,000 individuals living in Malaysia have epilepsy [3]. According to the World Health Organization, epilepsy is one of the most common neurological diseases [4]. Patients with epilepsy (PWEs) face many challenges related to their disease as well as to their families and health care providers (HCPs) [5]; additionally, lack of disease control affects the patients' work, relationship, and daily responsibilities [6]. HCPs also find it hard to monitor their PWEs to understand if any of their symptoms are improving or if they are experiencing any new symptoms [7]. In Malaysia, based on our experience, most PWEs use the calendar for seizure monitoring and some also record their medications on it.

PWEs are encouraged to self-manage their condition, as this has been proven to improve quality of life and considers the priority of their care [8]. However, the features of existing apps vary from one another, making it difficult to recommend one that is most complete to fulfill the needs of PWEs.

Self-management (SM) is generally defined as an individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle modifications necessary when living with a chronic disease [9]. SM is a dynamic process, which includes steps and decisions taken daily in response to living with a chronic condition [9]. In this process, patients are empowered for changing their health behavior via better knowledge about their disease and treatment, and by managing symptoms and physical/psychosocial consequences of the disease [10]. Patients need to rely on their self to manage their disorder daily [11].

The Institute of Medicine advocates the development of epilepsy self-management (eSM) programs to ensure the patients benefit from technologies such as computers and mobile devices [12]. Mobile health (mHealth) apps, including eSM apps, present daily companions of health data which help patients and health care professionals improve health outcomes [13] through constant monitoring of the patient's status.

mHealth apps play a major role among patients with chronic disease in managing their disorders [14]. According to one report, there were 325,000 mHealth apps on the market in 2017 [15]. There are about 165,000 mHealth apps in major app stores and most of these relate to management of diseases [16].

Many studies have found that mobile apps can assist patients with chronic condition [17]. Recently, a multitude of mHealth apps have been launched and shown to improve chronic disease care in patients with diabetes mellitus, hypertension, and other disorders [12,14,18].

The use of smartphone apps could be a promising strategy for seizure SM [14]. Nearly 94.4% of PWEs owned a smartphone; thus, it could be said that there is a positive attitude toward using epilepsy apps among PWEs [14].

This study aimed to perform a systematic review to identify the free eSM apps available in Apple App Store and Google Play app stores and to identify the domains covered by these eSM apps.

Methods

Study Overview

In this study we performed a systematic review of all apps that relate to PWEs and help these patients manage their disorder. We searched for eSM apps in both iOS (App Store) and Android (Google Play) app stores.

Research Questions

This study aimed to answer the following 3 questions:

- Research Question 1: What are the apps available to help in eSM?
- Research Question 2: What are the features of these apps?
- Research Question 3: What is the user rating of mobile apps related to eSM?

Inclusion Criteria

We performed a search of apps on Google Play and App Store using the keywords "epilepsy" or "seizures" from May to August 2018. The research was conducted in Malaysia; therefore, we included only Android and iOS apps, as only these are available in the country.

Inclusion criteria for choosing mobile apps were as follows: available in English, can be freely downloaded, and intended for individuals with epilepsy, covering at least one of the 11 domains described by Escoffery et al [9], which include health care communication, social support, medication adherence, treatment management, seizure response, wellness, stress management, coping and stress management, seizure tracking, safety, and proactivity.

Exclusion Criteria

Apps with installation issues or in other languages were excluded.

Review Process

To assess the quality of the apps and risk of bias, we first read and reviewed the app's description on the download page, to ascertain that the app fulfilled the inclusion or exclusion criteria. However, the app description alone is not enough to assess the app's quality [19], and therefore all selected apps were downloaded for further analysis [19]. A user's rating scale (score range 0-5) was used to determine the quality of the apps, with 5 indicating the best quality.

Results

Search Strategy and App Characteristics

A total of 382 mobile apps were found in the 2 app stores: 241 for Android and 141 for iOS. Apps that were not in English and not free were excluded (n=104), thus eventually 278 apps were downloaded for analysis. Next, we reviewed the description of each app to determine its features and purposes and excluded 235 apps, such as conference (meeting), games, and flashlights, which were not related to eSM. Eventually, 43 apps fulfilled

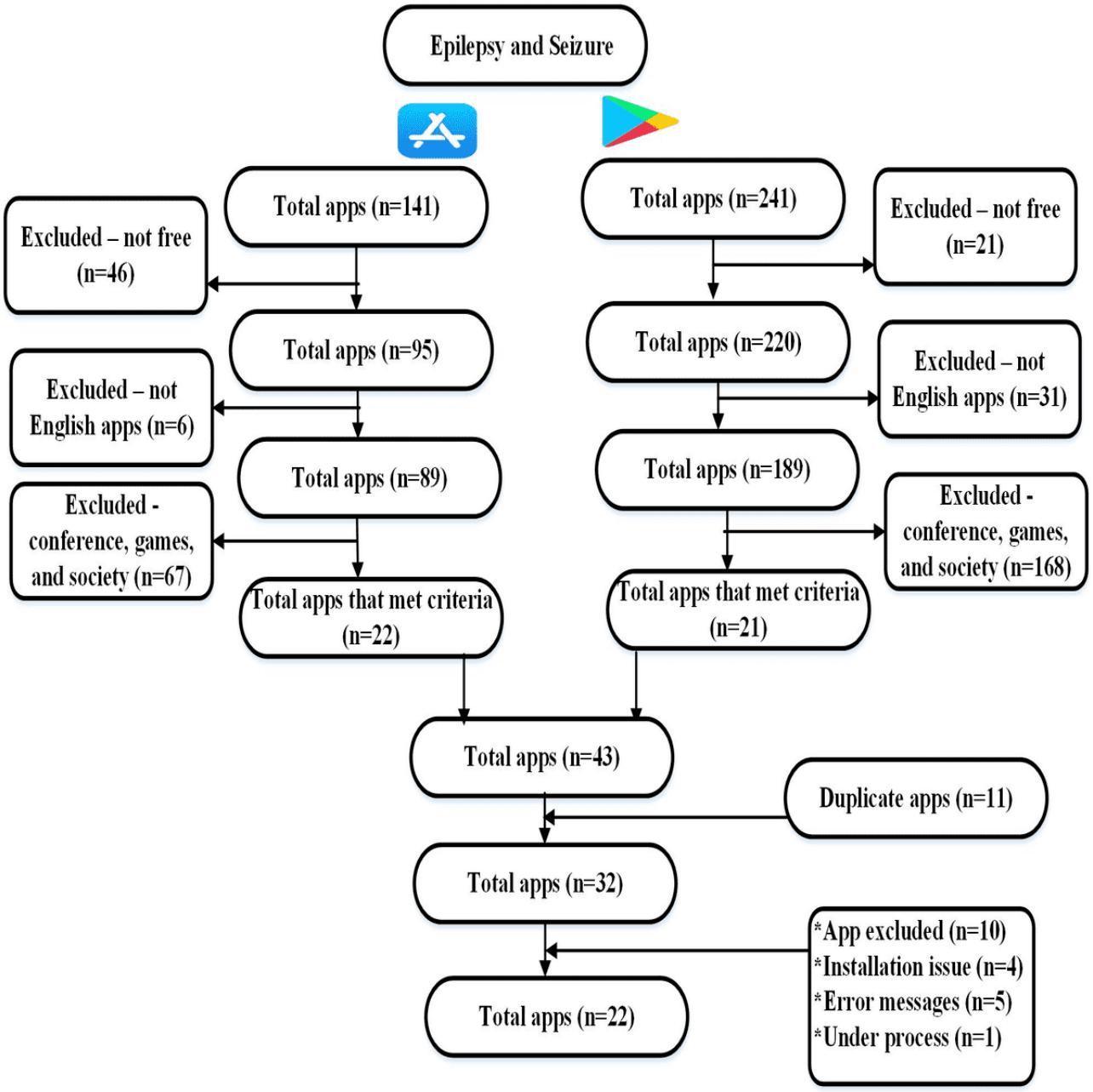
our criteria covering 1 or more eSM domains described by [9], of which 11 were compatible with both Android and iOS. We included 32 eSM apps for further analysis; however, 10 of these apps were excluded because of installation issues or they were malfunctioning. Table 1 presents the reasons for excluding the 10 apps during the installing stage. Of the final 22 included for analysis, 6 (27%) run on iOS, 7 (32%) on Android, and 9 (41%) on both operating systems. A PRISMA flow diagram showing the app selection process is presented in Figure 1. All 22 apps enhance eSM; however, none of the apps indicated whether any of the stakeholders, such as physicians, pharmacists, and PWEs, had participated in the app development process.

Most apps are generally complicated and the PWEs has to learn how to enter their data on the app, thereby affecting the potential use of these apps. Patients with chronic diseases such as epilepsy need a user-friendly app. As it is imperative that the content and features of epilepsy support apps motivate and support PWEs to manage their illness, PWEs and HCPs should be given an opportunity to participate and collaborate in developing a mobile app for eSM. The home page for each app contains a brief description about the app to convince the users to download it. This is especially the case with paid apps [19]. However, some app descriptions do not reflect the actual features and functions of the app. None of the apps indicated on their home page whether they have managed to achieve their goals through acceptance by the users. There was also no statement about the participation of stakeholders while developing the apps.

Table 1. Epilepsy self-management apps with issues during the installation stage.

App name	Operating system	Issue
Total Epilepsy Recorder	iOS	Error message
Seizure Sync Epilepsy Log	iOS	Error message
Fable Epilepsy Charity	iOS	App has been temporarily disabled by publisher
Wacean	iOS	Error message
Epilepsy Tracl	Android	This is currently a work in progress
Seizure Tracker	Android	Error message
Seizure Sync Epilepsy Manager	Android	Error message
Total Epilepsy	Android	Installation issue
Alert	Android	Installation issue
Dr. Epilepsy	Android	Installation issue

Figure 1. PRISMA flow diagram.



User's Rating

We used the apps' home page to obtain the user ratings. Table 2 presents the average user rating for the apps.

The ratings shown in Table 2 also depict the popularity of certain apps, as rated by the users. An app is usually considered popular if the users are happy with it, which eventually provides an indication of its usefulness.

Table 2. App user rating.

Name (app developer)	Number of installations	Average user rating	Number of raters
Epilepsy Journal (Olly Tree Application)	10,000+	4.4	176
Epilepsy Tool Kit (Epilepsy Society)	10,000+	3.9	79
EpiDiary (EpiDiary Irody, Inc.)	10,000+	3.5	345
Young Epilepsy (Synergix Health Ltd.)	5000+	4.1	88
Seizure Log (Seizure Tracker LLC)	5000+	4.7	9
Simple Seizure (NA[a])	5000+	4.4	176
My Seizure Diary (Epilepsy Foundation)	5000+	2.2	52
Epilepsy Ireland Diary App (DXC Technology)	1000+	3.1	11
Seizure First Aide (Afixia LLC)	1,000+	3.4	12
Win Over Epilepsy (WOE) (Bee Mobile Pvt. Ltd.)	500+	4.6	19
Birdhouse – for Epilepsy (Birdhouse LLC)	500+	3.0	2
EpApp (PENNSW & The Sydney Children's Hospitals Network)	100+	3.8	8
BioMark Health Epilepsy (NA)	100+	4.7	12
E-Action Info (The ImageFactory)	100+	4.00	2
E-Epilepsy Inclusion (The Hong Kong Society for Rehabilitation)	100+	4.9	7
ElFY Epilepsy (ELFY APP)	10+	5.00	2
Helpilepsy – Epilepsy Assistant & Diary (M4KEIT)	500+	4.8	8
Epilepsy Help (Dr. Bindu Menon Foundations)	—[b]	4.8	16
Epilepsy Safe (Comalf)	—	3.7	6
Epi & Me 2 (HandMe)	—	0	—
M.E. (Epilepsy Foundation of New Jersey)	—	0	—
SAMi3 Sleep Activity Monitor (HiPass Design LIC)	—	—	0

[a]NA: not available.

[b]—: Not applicable

Domains of Self-Management

Apps analyzed covered a total of 11 domains. The common domains were seizure tracking and seizure response, followed by treatment management and medication adherence. Three apps (Epilepsy Journal, Epilepsy Tool Kit, and EpiDiary) were installed more than 10,000 times, with their features specifically focusing on a few domains (treatment management, medication adherence, health care communication, and seizure tracking). Two apps covered over 6 SM domains (Young Epilepsy and E-Epilepsy Inclusion); however, both had lower installation rates (5000+ and 100+, respectively). Apps available only in Google Play or both Google Play and App Store had higher installation rates than the those available in App Store only (n=6), which include Seizure First Aide, E-Epilepsy Inclusion, ElFY Epilepsy, Epi & Me 2, M.E., and SAMi3 Sleep Activity Monitor (Tables 2 and 3).

Table 3. Name of the apps identified in this study and their eSM features (n=22).

App name (developer)	Operating system	Self-management features
Epilepsy Journal (Olly Tree Application)	Android	Helps to track seizure frequency; identify the effectiveness of epilepsy treatments; and input data on symptoms related to epilepsy that help HCPs to adjust medication dose or change the medication
Epilepsy Tool Kit (Epilepsy Society)	Both	Uses a new and interactive way by which it helps PWEs to manage their epilepsy
EpiDiary (EpiDiary Irody, Inc.)	Android	Keeps track of seizures, medicines, sleep, and how one feels daily
Young Epilepsy (Synergix Health Ltd.)	Both	Useful for young people with epilepsy and for parents or caregivers of a child who has epilepsy. Provides up-to-date information and has portal, video, and diary functionalities that help track seizures and manage symptoms
Seizure Log (Seizure Tracker LLC)	Both	Records seizure attacks
Simple Seizure (NA[a])	Android	Used as a diary to manage epilepsy
My Seizure Diary (Epilepsy Foundation)	Both	Epilepsy management includes self-monitoring and tracking, managing medications, and communicating with health care providers
Epilepsy Ireland Diary App (DXC Technology)	Both	Used to track and record seizures and also determine trigger factors
Seizure First Aide (Afixia LLC)	iOS	Educates about epilepsy
Win Over Epilepsy (WOE) (Bee Mobile Pvt. Ltd.)	Android	Helps to keep track of seizures, medicines, visits, etc.
Birdhouse – for Epilepsy (Birdhouse LLC)	Both	Helps to identify seizure triggers, manage a medication log, evaluate diets, and ensure the path to success
EpApp (PENNSW & The Sydney Children's Hospitals Network)	Both	Provides information about epilepsy; also has a tool that supports self-management for PWEs and their parents
BioMark Health Epilepsy (NA)	Android	Offers multiple functionalities (eg, passive data collection, medication reminders, quantified self-data collection, and care-team management) to understand epilepsy better and achieve more peace of mind for caregivers
E-Action Info (The ImageFactory)	Both	Teaches and increases the knowledge on epilepsy in a fun and easy way
E-Epilepsy Inclusion (The Hong Kong Society for Rehabilitation)	iOS	Assists in enhancing epilepsy self-management; helps to communicate with health care providers; and educates the public about epilepsy
ElFY Epilepsy (ELFY APP)	iOS	Tracks the right medication every day
Helpilepsy – Epilepsy Assistant & Diary (M4KEIT)	Both	Assists, tracks, and shares epilepsy records more easily and effectively
Epilepsy Help (Dr. Bindu Menon Foundations)	Android	Helps patients with epilepsy to organize and upload their medical data, keep medicine alarm and seizure alert, track appointment schedule, and know about epilepsy
Epilepsy Safe (Comalf)	Android	Helps to receive emergency aid from passersby in case of seizure attack in a public place
Epi & Me 2 (HandMe)	iOS	Assists in epilepsy management such as collecting and storing data on seizures, medication adherence, and life circumstances
M.E. (Epilepsy Foundation of New Jersey)	iOS	Useful for people with epilepsy and their families to manage their epilepsy
SAMi3 Sleep Activity Monitor (HiPass Design LIC)	iOS	Monitors the sleep activity of the patient for the caregiver and family members to carefully observe for any abnormal movement at night

[a]NA: not available.

Mobile App Description

This section describes the 22 apps which have been identified through this systematic review. Table 3 presents the name of the app and the self-management features included, while the eSM domain(s) they cover are listed in Table 4.

Table 4. App names and domains of self-management.

App name	Treatment management	Medication adherence	Health care communication	Seizure tracking	Seizure response	Safety	Wellness	Social support	Coping	Stress management	Proactivity	Total domains
Epilepsy Journal	✓	✓	✓	✓								4
Epilepsy Tool Kit	✓	✓	✓	✓	✓	✓						6
My Seizure Diary	✓	✓	✓	✓	✓			✓				6
EpApp	✓		✓	✓	✓							4
Young Epilepsy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		10
Helpilepsy – Epilepsy Assistant & Diary	✓	✓		✓	✓							4
EpiDiary	✓	✓	✓	✓								4
Simple Seizure			✓	✓	✓							3
Epilepsy Help	✓	✓		✓	✓	✓					✓	6
BioMark Health Epilepsy	✓	✓		✓	✓							4
Epilepsy Ireland Diary App			✓	✓	✓							3
Epilepsy Safe	✓	✓	✓	✓				✓				5
Win Over Epilepsy (WOE)	✓	✓	✓	✓	✓			✓				6
Birdhouse – for Epilepsy	✓	✓	✓	✓	✓	✓						6
E-Action Info	✓	✓	✓	✓	✓	✓						6
E-Epilepsy Inclusion	✓	✓	✓	✓	✓	✓			✓	✓	✓	9
ElFY Epilepsy	✓	✓	✓	✓	✓		✓					6
Seizure Log	✓		✓	✓	✓							4
Epi & Me 2	✓	✓		✓	✓							4
Seizure First Aide				✓	✓	✓						3
M.E.				✓	✓							2
SAMi3 Sleep Activity Monitor				✓	✓	✓						3

Treatment Management

A total of 17/22 apps (77%; eg, Epilepsy Journal, Epilepsy Tool Kit, Young Epilepsy, Epilepsy Help, Epi & Me 2) help patients self-manage their treatment. The common features in these apps were a reminder for treatment, appointment, and recommendations and advice of HCPs that should be followed.

Medication Adherence

A total of 15 apps, such as the Epilepsy Journal, My Seizure Diary, Birdhouse – for Epilepsy, EpiDiary, BioMark Health Epilepsy, Win Over Epilepsy (WOE), ElFY Epilepsy, and Seizure Log, were found to have features related to medication adherence. These apps enabled PWEs to establish and record a medication list, including the number of repeats remaining, the personalized dose, how many days the supply would last, and an alert reminder. Moreover, they provide PWEs a chance to register their time of medication, type of medication received, and frequency of acquiring the medication with their medication photo attached.

Health Care Communication

The success of the patient-centered model depends on communication between HCPs and patients [20], which strengthens their relationship and facilitates better treatment. In our study, 16 apps (73%), such as E-Epilepsy Inclusion, Simple Seizure Diary, Epilepsy Tool Kit, and Epi & Me, covered health care communication. These apps allow the PWEs to send their reports to their HCPs through email or share it as an efile during an appointment. This feature helps patients, caregivers, and HCPs to establish a relationship, which is extremely important. In addition, some apps allow the patient and HCPs to monitor the side effects of epilepsy, such as depression, anxiety, unstable emotions, and sleep problems.

Seizure Tracking

Seizure tracking provides critical information that helps the health care professional to evaluate the medication and health treatment received by the patients. All the apps cover seizure tracking and provide the features of tracking, such as registration of seizure time, duration, type and frequency of seizure, seizure triggers, and video recordings of the seizure.

Seizure Response

In our study, 19 apps (86%), such as Young Epilepsy, Seizure First Aide, Help Epilepsy, Seizure Log, BioMark Health Epilepsy, Epilepsy Tool Kit, Epilepsy Ireland Diary App, and Epilepsy Safe, cover the domain of seizure response. These apps include features such as first-aid instructions on how one can help someone when they have a seizure and when to dial 999. A customized emergency SMS text message can be sent with a touch of a button during the aura stage or following a seizure. The user can automatically send an alert message (with GPS location) to his/her family through the Epilepsy Safe app.

Safety

Eight apps (36%), such as Birdhouse – for Epilepsy and E-Action Info, covered the safety domain, which relates to avoiding risks that are detrimental to PWEs. One app from the App Store provided information for patients regarding the course of action during a seizure attack.

Wellness

Three apps, namely, ElFY Epilepsy, Birdhouse – for Epilepsy, and Young Journal, covered the wellness domain (compatible with both iOS and Android) and provide the PWEs with some type of wellness-related information, such as diet and exercise. The Birdhouse – for Epilepsy app also tracks food consumption to help identify associations with seizure activity.

Social Support

Four apps, namely, My Seizure Diary, Young Epilepsy, Epilepsy Safe, and Win Over Epilepsy, covered the social support domain and contained some stories related to epilepsy. This domain includes sharing stories or experiences with PWEs and getting some advice regarding epilepsy.

Coping and Stress Management

Coping, stress management, and proactivity are less popular domains and are included in only 2 apps, namely, ElFY Epilepsy and Young Epilepsy. These cover topics on managing stress; doing something that reduces sadness; and encourage PWEs to continue hobbies, relaxation, or exercises that help prevent seizure.

Proactivity

Only 2 apps provided some information related to proactivity (Epilepsy Help and E-Epilepsy Inclusion), and included some advice that may avoid situations or things that might cause a seizure.

Discussion

Principal Findings

In this review, 22 SM apps for epilepsy were found. The apps with the highest installation frequency were Epilepsy Journal, Epilepsy Tool Kit, and EpiDiary (10,000+ for each), and these cover the 4 most important domains of eSM, namely, treatment management, medication adherence, health care communication, and seizure tracking. However, as nearly 50 million people are estimated to have epilepsy worldwide [1,2], these apps are underutilized.

Monitoring seizure episodes assists in increasing the possibility of capturing seizures and tracking the evolution of the disease [21]. Using paper diaries to collect data related to patient's symptoms is a challenging task. Thus, mobile apps have become a great tool that help PWEs to cope with and control their condition. These apps also aid PWEs to record their data and evaluate the effect of medication and treatment administered in consultation with their HCPs. A well-designed app provides a good opportunity for both PWEs and HCPs to manage epilepsy and select the best option together.

Apps with the highest download/installation frequency reflect increased acceptability by users, despite covering less domains. Most of these apps focus on seizure tracking because this provides valuable information that helps in the management of the disease [21]. A high percentage (60%-70%) of patients can become seizure free when they promptly take the effective medication. Adherence is defined by the World Health Organization as "the extent to which a person's behavior—taking medication, following a diet, and executing lifestyle changes—corresponds with the agreed recommendations from a HCP" [22].

However, finding an appropriate antiepileptic drug (AED) can be a long and challenging process because patients struggle to report their seizure frequency [23]. Providing seizure information remains critical to help neurologists in selecting the correct dose of AEDs for their patients; efficiency of AED in long-term treatments should also be evaluated [23].

Nonadherence is a critical issue for PWEs [24] and can lead to an increase in seizure frequency [25]. PWEs generally do not adhere to their medication, which negatively affects their situation [26]. Nonadherence to medications usually decreases the quality of treatment outcomes, maximizes the consultations and hospitalization, and increases the health care cost [27]. Despite the risks associated with not taking medications, 50% of the patients with a chronic condition fail to adhere to their treatment recommendations [22]. Adherence requires active involvement of the patient and a therapeutic alliance between the patient and his/her physician [28]. Therefore, most apps have features to support medication management.

According to our review, most of the apps cover at least one or more domains of eSM, as recommended by the World Health Organization [4]. Each app has features that allow PWEs to self-manage epilepsy. However, not all the apps analyzed covered all 11 domains. Every app focused on different domains of eSM. The Young Epilepsy app is the only one that covered

most (10 overall) eSM domains (user rating: 4.4), whereas the My Seizure Diary app covered only 6 domains and had the lowest user rating (2.2).

Without effective management, PWEs will have poor quality of life, and so mHealth apps are considered promising tools that will assist and fill the gap in eSM. End users should be part of design and development stages of apps to increase the quality of apps and achieve the app's intended goal(s).

Study Limitations

We only included English apps because this study aimed to review the current epilepsy apps from an international perspective, and not just from a Malaysian perspective. However, the apps were accessed through the Malaysian App Store and Google Play sites as stated in the "Methods" section. The research was conducted in Malaysia, and the apps that are available for iOS and Android in this country differ from those in the United States, while some are not available in our country, which possibly affects the overall analysis. The study included only free apps. Some with reasonable prices might be added to make the study more valuable. Many apps were excluded from the study because their languages were not English, although some of these have great features. Only Android- and iOS-compatible apps were targeted in this research. However, many related apps could be available for other mobile operating systems, such as BlackBerry or LiteOS (Huawei).

Conclusion and Future Work

A total of 22 apps were found to support different eSM domains; however, almost all of these were underutilized (ie, had less installation rate). Thus, to improve the efficacy and availability of these apps, we propose the following: (1) involve various stakeholders, such as physicians, pharmacists, and PWEs, during the development of mHealth apps; (2) assess the efficacy and acceptance of these apps objectively by performing a usability analysis; and (3) promote the apps so that they benefit more PWEs. Malaysia is a multiracial country which is dominated by 3 racial groups, namely, Malay, Indian, and Chinese. In the future, we will also analyze similar apps in local languages to ensure all important features are considered.

Acknowledgments

We thank Hussain Mutlaq Alnajrani, MS, University of Malaya, who assisted us with app installation to review the mobile apps. This study is self-funded.

Conflicts of Interest

None declared.

References

1. Amorim P, Moraes T, Fazanaro D, Silva J, Pedrini H. Electroencephalogram signal classification based on shearlet and contourlet transforms. Expert Systems with Applications 2017 Jan;67:140-147. [doi: 10.1016/j.eswa.2016.09.037]
2. Velasquez SE, Chaves-Carballo E, Nelson E. Pediatric Teleneurology: A Model of Epilepsy Care for Rural Populations. Pediatric neurology 2016 Nov;64:32-37. [doi: 10.1016/j.pediatrneurol.2016.08.001] [Medline: 27742162]
3. Norsa'adah B, Zainab J, Knight A. The quality of life of people with epilepsy at a tertiary referral centre in Malaysia. Health and quality of life outcomes 2013 Aug 23;11(1):1-6 [FREE Full text] [doi: 10.1186/1477-7525-11-143] [Medline: 23972031]
4. World Health Organization. Atlas: Epilepsy Care in the World. Geneva, Switzerland: World Health Organization; 2005.
5. England MJ, Liverman CT, Schultz AM, Strawbridge LM. Epilepsy across the spectrum: promoting health and understanding. A summary of the Institute of Medicine report. Epilepsy & Behavior 2012 Oct;25(2):266-276 [FREE Full text] [doi: 10.1016/j.yebeh.2012.06.016] [Medline: 23041175]
6. Dannenberg M, Mengoni SE, Gates B, Durand M. Self-management interventions for epilepsy in people with intellectual disabilities: A scoping review. Seizure 2016 Oct;41:16-25 [FREE Full text] [doi: 10.1016/j.seizure.2016.06.022] [Medline: 27447692]
7. Helmers SL, Kobau R, Sajatovic M, Jobst BC, Privitera M, Devinsky O, Centers for Disease ControlPrevention Managing Epilepsy Well Network. Self-management in epilepsy: Why and how you should incorporate self-management in your practice. Epilepsy & Behavior 2017 Mar;68:220-224 [FREE Full text] [doi: 10.1016/j.yebeh.2016.11.015] [Medline: 28202408]
8. Leenen LA, Wijnen BF, de Kinderen RJ, van Heugten CM, Evers SM, Majoie MH. Are people with epilepsy using eHealth-tools? Epilepsy & Behavior 2016 Nov;64(Pt A):268-272. [doi: 10.1016/j.yebeh.2016.08.007] [Medline: 27780086]
9. Escoffery C, Bamps Y, LaFrance WC, Stoll S, Shegog R, Buelow J, et al. Factor analyses of an Adult Epilepsy Self-Management Measurement Instrument (AESMMI). Epilepsy & Behavior 2015 Sep;50:184-189. [doi: 10.1016/j.yebeh.2015.07.026] [Medline: 26264465]
10. Leenen LA, Wijnen BF, de Kinderen RJ, Majoie MH, van Heugten CM, Evers SM. (Cost)-effectiveness of a multi-component intervention for adults with epilepsy: study protocol of a Dutch randomized controlled trial (ZMILE study). BMC neurology 2014 Dec 24;14(1):255 [FREE Full text] [doi: 10.1186/s12883-014-0255-3] [Medline: 25540089]

11. Hale K, Capra S, Bauer J. A Framework to Assist Health Professionals in Recommending High-Quality Apps for Supporting Chronic Disease Self-Management: Illustrative Assessment of Type 2 Diabetes Apps. JMIR Mhealth Uhealth 2015 Sep 14;3(3):e87 [FREE Full text] [doi: 10.2196/mhealth.4532] [Medline: 26369346]
12. Pandher PS, Bhullar KK. Smartphone applications for seizure management. Health informatics journal 2016 Jun 18;22(2):209-220 [FREE Full text] [doi: 10.1177/1460458214540906] [Medline: 25038202]
13. Bulaj G. Combining non-pharmacological treatments with pharmacotherapies for neurological disorders: a unique interface of the brain, drug-device, and intellectual property. Frontiers in neurology 2014 Jul 14;5:126 [FREE Full text] [doi: 10.3389/fneur.2014.00126] [Medline: 25071711]
14. Liu X, Wang R, Zhou D, Hong Z. Feasibility and acceptability of smartphone applications for seizure self-management in China: Questionnaire study among people with epilepsy. Epilepsy & Behavior 2016 Feb;55:57-61. [doi: 10.1016/j.yebeh.2015.11.024] [Medline: 26745631]
15. Paglialonga A, Schiavo M, Caiani EG. Automated Characterization of Mobile Health Apps' Features by Extracting Information From the Web: An Exploratory Study. Am J Audiol 2018 Nov 19;27(3S):482-492. [doi: 10.1044/2018_aja-imia3-18-0008]
16. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
17. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
18. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatric diabetes 2013 Jun 30;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
19. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemedicine and e-Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
20. Shegog R, Begley CE, Harding A, Dubinsky S, Goldsmith C, Hope O, et al. Description and feasibility of MINDSET: A clinic decision aid for epilepsy self-management. Epilepsy & Behavior 2013 Dec;29(3):527-536. [doi: 10.1016/j.yebeh.2013.09.023]
21. Gu Y, Cleeren E, Dan J, Claes K, Van Paesschen W, Van Huffel S, et al. Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy. Sensors 2017 Dec 23;18(1):29 [FREE Full text] [doi: 10.3390/s18010029] [Medline: 29295522]
22. Rich A, Brandes K, Mullan B, Hagger MS. Theory of planned behavior and adherence in chronic illness: a meta-analysis. Journal of behavioral medicine 2015 Aug 21;38(4):673-688. [doi: 10.1007/s10865-015-9644-3] [Medline: 25994095]
23. Bidwell J, Khuwatsamrit T, Askew B, Ehrenberg JA, Helmers S. Seizure reporting technologies for epilepsy treatment: A review of clinical information needs and supporting technologies. Seizure 2015 Nov;32:109-117 [FREE Full text] [doi: 10.1016/j.seizure.2015.09.006] [Medline: 26552573]
24. Lin C, Chen H, Pakpour AH. Correlation between adherence to antiepileptic drugs and quality of life in patients with epilepsy: A longitudinal study. Epilepsy & Behavior 2016 Oct;63:103-108. [doi: 10.1016/j.yebeh.2016.07.042] [Medline: 27588360]
25. Lua PL, Neni WS. Feasibility and acceptability of mobile epilepsy educational system (MEES) for people with epilepsy in Malaysia. TELEMEDICINE and e-HEALTH 2012 Dec;18(10):777-784 [FREE Full text] [doi: 10.1089/tmj.2012.0047] [Medline: 23078181]
26. Paschal AM, Rush SE, Sadler T. Factors associated with medication adherence in patients with epilepsy and recommendations for improvement. Epilepsy & Behavior 2014 Feb;31:346-350. [doi: 10.1016/j.yebeh.2013.10.002] [Medline: 24257314]
27. Demonceau J, Ruppar T, Kristanto P, Hughes DA, Fargher E, Kardas P, et al. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013 May 16;73(6):545-562 [FREE Full text] [doi: 10.1007/s40265-013-0041-3] [Medline: 23588595]
28. 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 internal medicine 2016 Mar 01;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]

Abbreviations

eSM: epilepsy self-management
HCP: health care provider
PWEs: patients with epilepsy
SM: self-management

Edited by L Buis; submitted 13.07.20; peer-reviewed by K Fitzner, AM Lunde Huseboe; comments to author 11.10.20; revised version received 06.01.21; accepted 24.04.21; published 28.05.21.

Please cite as:
Alzamanan MZ, Lim KS, Akmar Ismail M, Abdul Ghani N
Self-Management Apps for People With Epilepsy: Systematic Analysis
JMIR Mhealth Uhealth 2021;9(5):e22489
URL: https://mhealth.jmir.org/2021/5/e22489
doi:10.2196/22489
PMID:34047709

Original Paper

Postmarketing Safety Monitoring After Influenza Vaccination Using a Mobile Health App: Prospective Longitudinal Feasibility Study

Minh Tam H Nguyen[1], Staatsexamen; Gérard Krause[1,2], Prof Dr; Brigitte Keller-Stanislawski[3], Dr med; Stephan Glöckner[1], PhD; Dirk Mentzer[3*], Dr med; Jördis J Ott[1,2*], PD Dr

[1]PhD Programme Epidemiology, Hannover Biomedical Research School, Hannover Medical School, Hannover, Germany

[2]Hannover Medical School, Hannover, Germany

[3]Paul Ehrlich Institute, Langen, Germany

[*]these authors contributed equally

Corresponding Author:
Jördis J Ott, PD Dr
PhD Programme Epidemiology
Hannover Biomedical Research School
Hannover Medical School
Hannover
Germany
Phone: 49 531 6181 3115
Email: joerdis.ott@helmholtz-hzi.de

Abstract

Background: For the safety monitoring of vaccinations postlicensure, reports of adverse events after immunization (AEFIs) are crucial. New technologies such as digital mobile apps can be used as an active approach to capture these events. We therefore conducted a feasibility study among recipients of the influenza vaccination using an app for assessment of the reporting of AEFIs.

Objective: The goal of the research was to determine factors influencing adherence to and correct use of a newly developed app for individuals to report AEFI for 3 months using regular reminder functions, to identify determinants of AEFI occurrence and define reported AEFI types.

Methods: We developed the app (SafeVac) and offered it to recipients of the influenza vaccination in 3 occupational settings in fall 2018. In this prospective longitudinal feasibility study, data on AEFIs were generated through SafeVac for 3 months. Using logistic and Cox regression, we assessed associations between app adherence, correct app entry, AEFIs, and sociodemographic parameters.

Results: Of the individuals who logged into SafeVac, 61.4% (207/337) used the app throughout a 3-month period. App use adherence was negatively associated with female sex (odds ratio [OR] 0.47; CI 0.25-0.91) and correct app entry was negatively associated with older age (OR 0.96; CI 0.93-0.99) and lower education (OR 0.31; CI 0.13-0.76). AEFI occurrence was associated with female sex (hazard ratio 1.41; CI 1.01-1.96) and negatively with older age (hazard ratio 0.98; CI 0.97-0.99). The most common AEFIs reported were injection site pain (106/337), pain in extremity (103/337), and fatigue/asthenia (73/337).

Conclusions: Digital AEFI reporting was feasible with SafeVac and generated plausible results for this observation period and setting. Studies directly comparing SafeVac with conventional passive reporting schemes could determine whether such digital approaches improve completeness, timeliness, and sensitivity of vaccine vigilance. Further studies should evaluate if these results are transferable to other vaccinations and populations and if introduction of such a tool has an influence on vaccination readiness and therefore vaccine safety.

(JMIR Mhealth Uhealth 2021;9(5):e26289) doi:10.2196/26289

KEYWORDS

mHealth; mobile health; digital health; adverse event; adverse event following immunization; active reporting; pharmacovigilance; therapeutic use; adverse effect

Introduction

Vaccinations have been the most effective measure to prevent infectious diseases, preventing over 2 million deaths per year worldwide [1,2]. Especially for new and changing vaccines (eg, those against COVID-19 or the seasonal flu), intensified and timely surveillance of adverse events following immunization (AEFIs) might be of high relevance [3,4]. This is also indicated because these vaccinations are administered to many individuals with different characteristics and medical histories within a short period of time [4-6]. Although all licensed vaccines undergo clinical safety trials until they obtain regulatory approval, some AEFI become known only after marketing on a large scale [7,8]. In Germany, physicians are obliged by law to report suspicions of health impairments that exceed the normal degree of a vaccination to the local health authority (§6 Abs. 1, no. 3, IfSG), which is obliged to report to the Paul Ehrlich Institute (PEI), the Federal Institute for Vaccines and Biomedicines, via phone, mail, fax, or online. Vaccination recipients can also voluntarily contribute to vaccine safety via a web-based app [9]. While this spontaneous reporting system is a cost-effective way to detect adverse events, it is limited by imprecise information on the denominator, underreporting, and delay in reporting due to various media breaches in reporting [10-12]. A new way to capture AEFIs and overcome these limitations is to have patients send reports directly using a mobile app. Although more than 318,000 health apps were available in the app stores in 2017, the apps were primarily developed for exercises and fitness [13] rather than to capture adverse events. In fact, in a systematic review by Cashman et al [14], only a single participant-centered app capturing AEFIs in near real time was found [15]. This clearly shows a general lack of research on an app to report AEFIs, especially in long-term use. The long period of monitoring is particularly important to capture previously unknown unexpected AEFIs with late onset. Furthermore, in epidemiological studies, longitudinal data can help to assess causal relationships and explore determinants [16].

Therefore, our aim was to (1) assess whether an app for reporting AEFIs can be used for 3 months, (2) determine factors influencing the adherence to and correct use of a newly developed app (SafeVac, Paul Ehrlich Institute, and Helmholtz Centre for Infection Research [HZI]) for individuals to report AEFIs for 3 months using regular reminder functions, and (3) identify determinants for AEFI occurrence and define reported AEFI types.

Methods

Participant Recruitment

We recruited participants for our prospective longitudinal study from staff of 3 different employers in Germany during fall 2018: Investitionsbank Berlin, PEI, and University Hospital Frankfurt. We included volunteers on 2 respective days. To be eligible, individuals had to be employed at these institutions, vaccinated against influenza by the occupational health physician on these days, own a smartphone, be at least 18 years old, and be proficient in the German language.

We gave participants who agreed to take part in our study an information leaflet comprising information about the study, data transmission, and data storage; a random ID; and instructions on downloading and using the SafeVac app. Once the individuals agreed to participate and signed the declaration of consent, we asked participants to download and log in to SafeVac immediately after receiving the influenza vaccination. In the app, participants provided sociodemographic information and their vaccination history and then entered information about occurrence or nonoccurrence of adverse events. After participants used the app, we distributed usability questionnaires based on the System Usability Scale (SUS) and evaluated them using the adjective rating scale [17,18]. Additionally, participants could take part in a lottery to win a tablet or smartwatch. This was offered to all participants who completed the AEFI reports for at least 1 week.

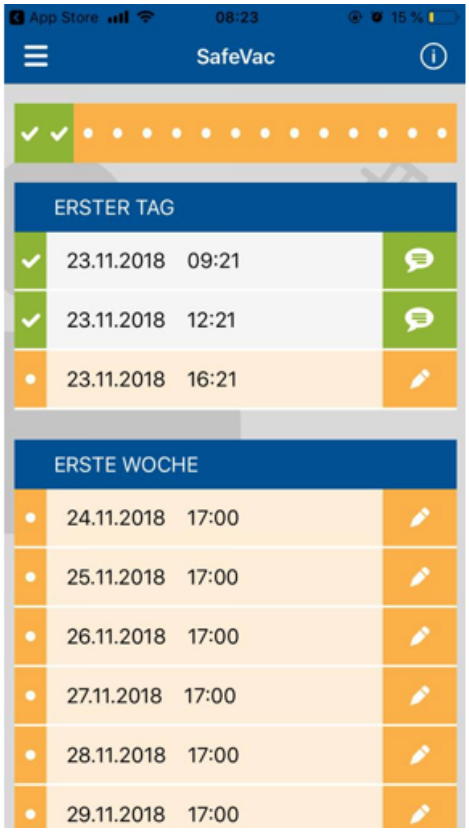
The study protocol and data protection concept were reviewed and assessed without any concerns by the ethical committee of the Medical Association of Lower Saxony in Germany, the HZI institutional data protection officer, and the Federal Commissioner for Data Protection and Freedom of Information.

AEFI Reporting and Conception of the SafeVac App

In SafeVac, we asked participants to enter information on receipt of influenza and previous vaccinations and sociodemographic variables and describe any AEFI occurrence. We asked these questions at 15 intervals defined over a period of 3 months: 1 hour, 4 hours, and 8 hours after vaccination; daily until the 7th day after vaccination; weekly for 3 weeks after vaccination; and monthly for 2 subsequent months after vaccination. In case of AEFI occurrence, participants answered additional questions in the app based on requirements of the online reporting platform for adverse events hosted by the national responsible authorities PEI and the Federal Institute for Drugs and Medical Devices. Participants could select a specific AEFI (eg, fatigue) from a drop-down list or enter it manually in the comment text field. In addition, participants were asked to provide additional information about chronic medical conditions, medications taken for chronic conditions, and pregnancy status.

SafeVac was available in Android (version 4.4 or higher) and iOS (10 or higher) app stores. For the development and design of SafeVac, we took a user-centered approach for which we gathered data in a previous study on users' app preferences [19]. Based on these results, we used gamification elements such as the appearance of puzzle pieces and a loading bar to enhance app use adherence (Figure 1). Every app entry was transmitted anonymously and securely to PEI using https/SSL encryption. To evaluate the feasibility of the app, we used the adherence rate as a surrogate parameter and assessed whether AEFIs entered into the app were in accordance with the literature.

Figure 1. Screenshot of SafeVac app.



Data Management and Data Analysis

We included all information collected through SafeVac from fall 2018 until March 14, 2019, in our study. Before analyzing the data, we excluded test IDs. We coded AEFI entries in the free-text fields according to the preferred terms in the Medical Dictionary of Regulatory Activities and validated them using the four-eyes principle. We also checked for consistency in cases where participants had entered AEFIs manually in the comment text field and additionally selected them from the drop-down list at the same time point. In case of discrepancy, we coded them as 2 different types of AEFIs. If participants entered 2 identical AEFIs for the same time point, we removed one. If the free-text field described the AEFI in more detail than the AEFI selected on the drop-down list, we ignored the selected AEFI and coded the manually entered AEFI systematically according to the corresponding preferred term (eg, if "pain in extremities" was selected in the drop-down list and "pain in the arm (injection site)" was entered manually, we coded the AEFI as "injection site pain"). The BMI was calculated and categorized using the World Health Organization scale approach [20].

By using logistic regression, we estimated the relationship between sociodemographic variables and vaccination uptakes with the binary outcomes app adherence until the end of the study, AEFI occurrence and correct entry of vaccination information, respectively. We defined incorrect entry of vaccination information as any misspellings in the name of the received influenza vaccination or its associated batch number but did not take case sensitivity into account. To find determinants for the outcome AEFI occurrence in 3 months, we used a Cox regression model. Variables were selected by using a backward selection by Akaike information criterion. In all models we set age and sex as a priori confounders. Missing data were not included in the model.

Additionally, we used Cox regression to explore determinants of reporting an influenza-like illness as an AEFI. We used R software version 3.2.5 (R Foundation for Statistical Computing) for data analysis and visualization.

Study Participation and Evaluation of SafeVac

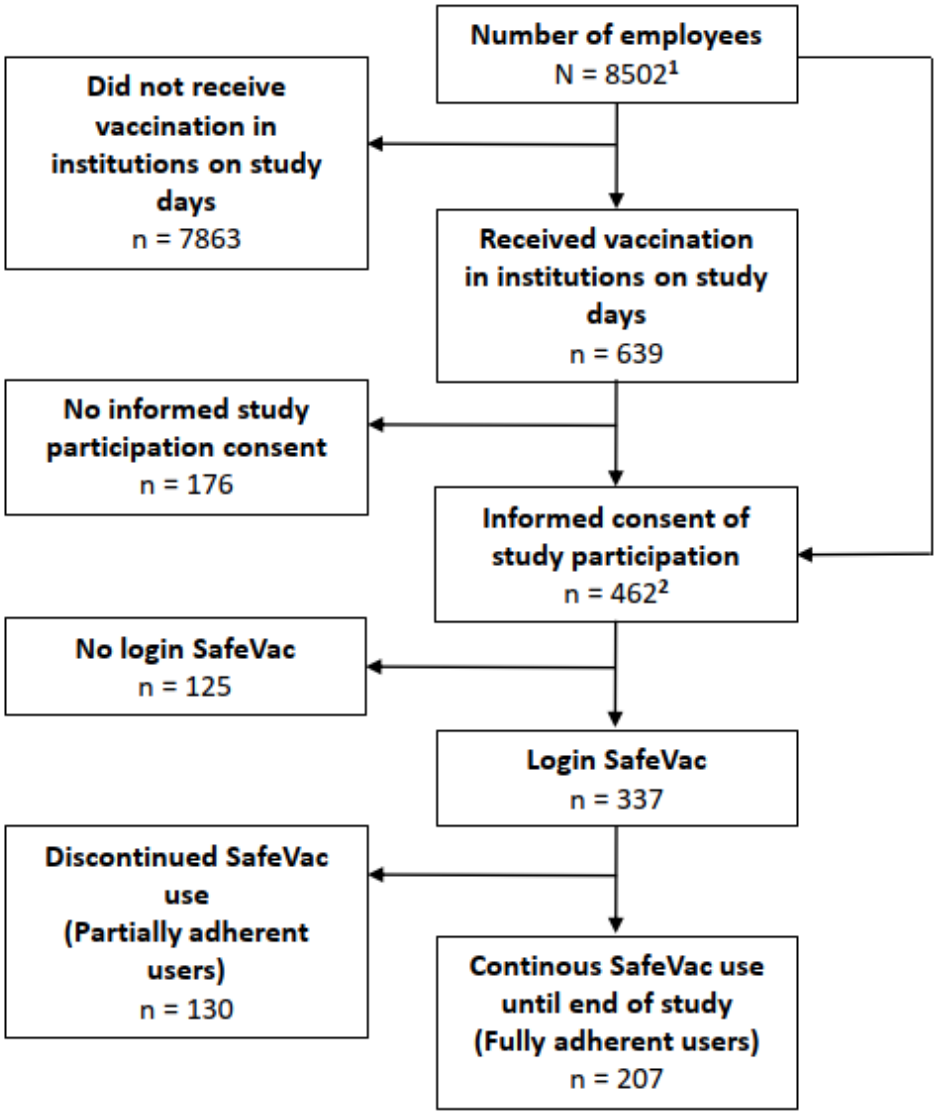
To assess quality and usability of the app, we developed a feedback questionnaire in LimeSurvey, an online survey tool, by including the SUS and using an approach of the User Version of the Mobile Application Rating Scale [21,22]. We distributed the feedback questionnaire through a link 1 year after the start of the study via the internal websites of the 3 participating institutions.

Results

Participant Characteristics and AEFI Reporting Over Three Months

Of the participants who provided informed consent, 72.9% (337/462) logged into the SafeVac (study population) and made a minimum of one app entry; 61.4% (207/337) used the app until the end of the study (Figure 2).

Figure 2. Study participant recruitment. [1]Obtained through personal communication and estimates based on institutional website; for PEI and IBB only ≥18 years old included; [2]included participants who were vaccinated before study days.



The majority of participants were female (224/337, 66.5%) and 84.0% (283/337) had a general certificate of education (Abitur). Almost half of the participants (166/337, 49.3%) stated that they had also been vaccinated against influenza in an earlier year. We assessed characteristics according to adherence of app use in the study (fully vs partially adherent participants), and they appear mostly similar (eg, 74/207 [35.7%] vs 39/130 [30.0%] male participants; Table 1).

Table 1. Characteristics of study participants according to adherence to app use.

Characteristic	Fully adherent participants[a] (n=207)	Partially adherent participants[b] (n=130)	Total (n=337)
Age in years, mean	35.7	37.3	36.3
Sex, n (%)			
Male	74 (35.7)	39 (30.0)	113 (33.5)
Female	133 (64.3)	91 (70.0)	224 (66.5)
General certificate of education, n (%)			
Yes	172 (83.1)	111 (85.4)	283 (84.0)
No	31 (15.0)	17 (13.1)	48 (14.2)
Missing	4 (1.9)	2 (1.5)	6 (1.8)
Vaccination place, n (%)			
FFM[c] (Frankfurt Main)	150 (72.5)	75 (57.7)	225 (66,8)
PEI[d] (Langen)	23 (11.1)	22 (16.9)	45 (13.4)
IBB[e] (Berlin)	30 (14.5)	30 (23.1)	60 (17.8)
Missing	4 (1.9)	3 (2.3)	7 (2.1)
Vaccinated against influenza last year, n (%)			
Yes	103 (49.8)	59 (45.4)	162 (48.1)
No	28 (13.5)	24 (18.5)	52 (15.4)
Missing	76 (36.7)	47 (36.2)	123 (36.5)
Number of influenza vaccinations within the last 5 years, n (%)			
0	58 (28.0)	37 (28.5)	95 (28.2)
1	38 (18.4)	13 (10.0)	51 (15.1)
2	23 (11.1)	15 (11.5)	38 (11.3)
3	24 (11.6x)	19 (14.6)	43 (12.8)
4	21 (10.1)	16 (12.3)	37 (11.0)
5	37 (17.9)	24 (18.5)	61 (18.1)
Cannot remember	6 (29.0)	5 (3.8)	11 (3.3)
Missing	0	1 (0.8)	1 (0.3)

[a]Fully adherent participants defined as participants who replied to questions on adverse event following immunization at all app notification time points.

[b]Partially adherent participants defined as participants who replied to questions on adverse event following immunization to some app notification time points.

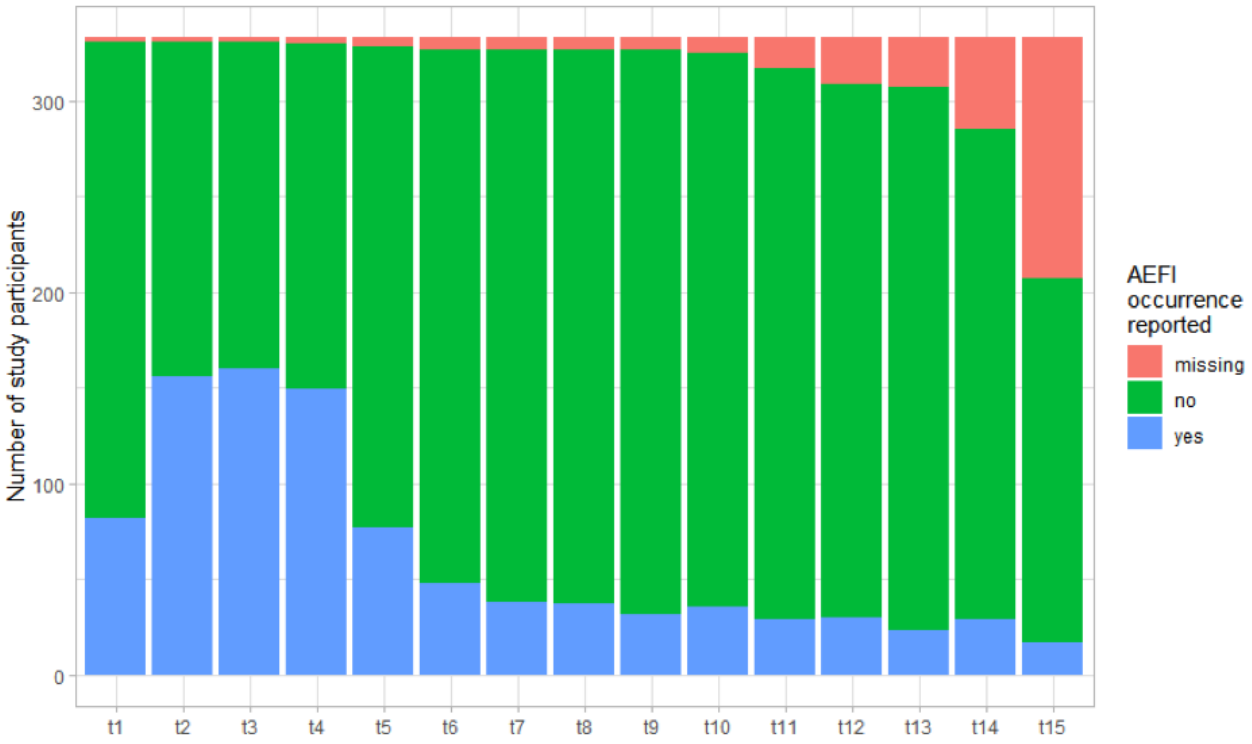
[c]FFM: University Hospital Frankfurt.

[d]PEI: Paul Ehrlich Institute.

[e]IBB: Investitionsbank Berlin.

Of the participants who logged in, 81.4% (271/333) reported experiencing one or more AEFIs after vaccination. We found a rise in reported AEFIs at 4 hours after vaccination (156/333, 47.8%); thereafter, the number of participants reporting an AEFI remained steady over 8 hours (160/333, 48.0%) and 1 day (149/333, 44.7%) after vaccination. A decline started from 2 days after vaccination (77/333, 23.1%) onward. Study participant reporting attrition increased slightly over time, with the highest attrition difference (78) between 56 days and 84 days after vaccination (Figure 3).

Figure 3. Reported adverse event following immunization occurrences per time point after influenza vaccination.

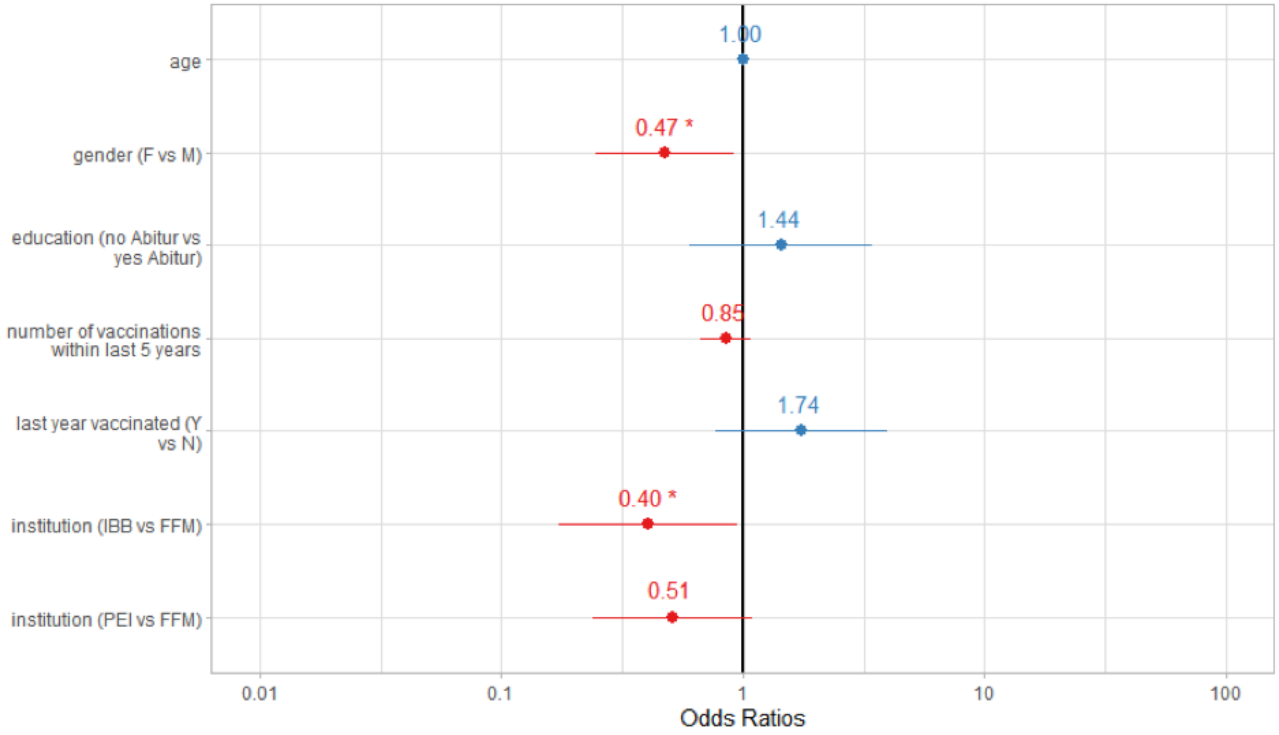


Time points of push notifications:
t0=time point of vaccination,
t1=t0+1h, t2=t0+4h, t3=t0+8h,
t4= t0+1d, t5= t0+2d, t6= t0+3d, t7= t0+4d, t8= t0+5d, t9= t0+6d, t10= t0+7d,
p11= p0+14d, p12= p0+21d, p13= p0+28d, p14= p0+56d, p15= p0+84d
4 were excluded due to divergent time intervals

Factors Associated With Adherence to App Use

Regarding adherence to SafeVac use until the end of follow-up, our logistic regression revealed a decrease of 43% (odds ratio [OR] 0.47; CI 0.25-0.91) for females and staff of the banking institution. The latter population was significantly less adherent than staff of the university hospital (OR 0.40; CI 0.17-0.94) after adjusting for age, sex, education, influenza vaccination received in the last year, number of influenza vaccinations in previous 5 years, occurrences of AEFIs, and institutional affiliation (Figure 4).

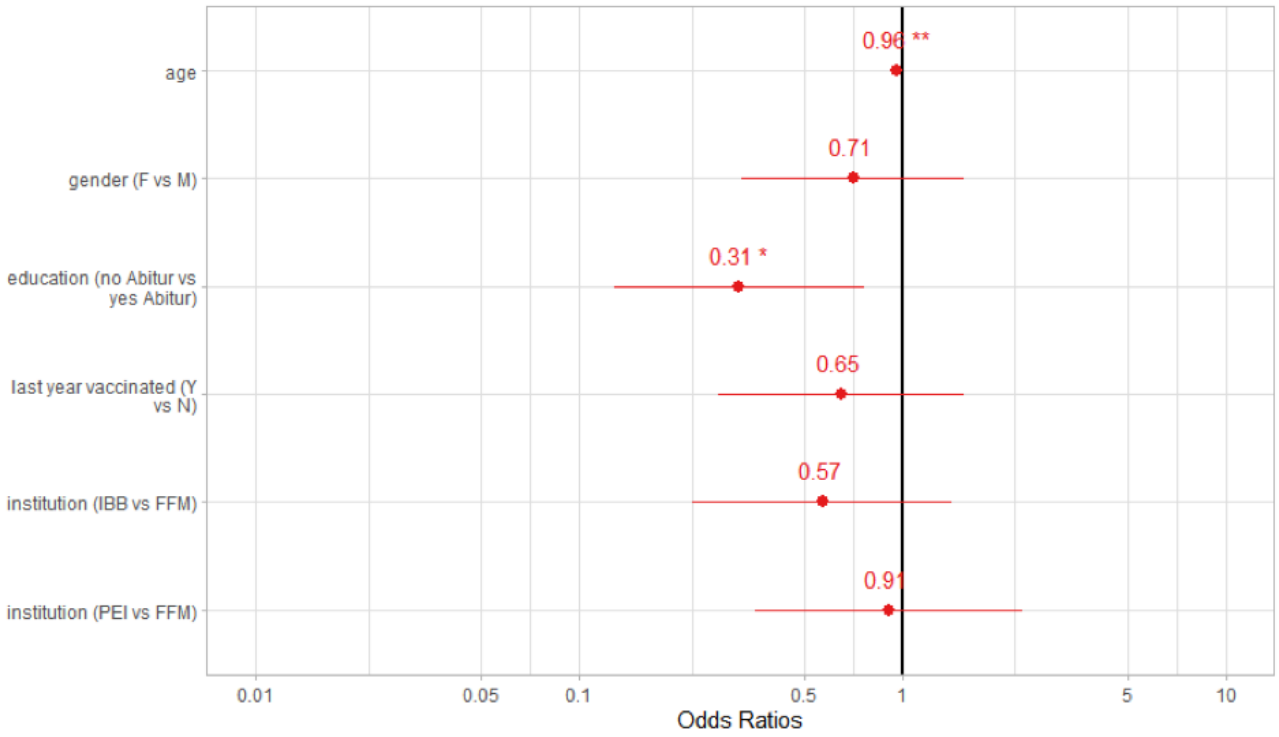
Figure 4. Factors associated with fully adherent app use versus partially adherent app use over 3 months, pseudoR2 (McFadden) = 0.065. F: female; M; male; Y: yes; N: no; IBB: Investitionsbank Berlin; FFM: University Hospital Frankfurt; PEI: Paul Ehrlich Institute.



Factors Associated With Correct App Entry of Vaccination

Overall, 260 participants correctly entered the data for their received vaccination. Having no Abitur (OR 0.31; CI 0.13-0.76) and increasing age (OR 0.96; CI 0.93-0.99) were negatively associated with correctness of entry of vaccine information (Figure 5). Other factors we analyzed (eg, age, gender, vaccination received last year, and affiliated institution) were not significantly associated with correct app entry of vaccination.

Figure 5. Factors associated with correct versus incorrect entry of vaccination information, pseudoR2 (McFadden) = 0.124. F: female; M; male; Y: yes; N: no; IBB: Investitionsbank Berlin; FFM: University Hospital Frankfurt; PEI: Paul Ehrlich Institute.



Most Reported AEFI Occurrences From Study Participants

The most mentioned AEFIs from all participants were injection site pain (106/337, 31.5%), followed by pain in extremity (103/337, 30.6%) and fatigue/asthenia (73/337, 21.7%). Proportional differences in reporting were found in fully adherent and partially adherent participants (eg, pain in extremity, 60/207 [29.0%], 43/130 [33.1%]). However, there were no statistically significant differences in AEFI reporting between fully adherent and partially adherent app users (Table 2).

Table 2. Most common AEFIs according to app use adherence.

Adverse event	Fully adherent participants reporting AEFI[a] (n=207), n (%)	Partially adherent participants reporting AEFI (n=130), n (%)	*P* value[b]	Participants reporting AEFI (n=337), n (%)
Injection site pain	66 (31.9)	40 (30.8)	.83	106 (31.5)
Pain in extremity	60 (29.0)	43 (33.1)	.42	103 (30.6)
Fatigue/asthenia	42 (20.3)	31 (23.8)	.44	73 (21.7)
Headache	39 (18.8)	23 (17.7)	.79	62 (18.4)
Influenza-like illness	38 (18.4)	18 (13.8)	.27	56 (16.6)
Myalgia	39 (18.8)	17 (13.1)	.17	56 (16.6)
Rhinitis	31 (15.0)	20 (15.4)	.91	51 (15.1)
Throat irritation	25 (12.1)	21 (16.2)	.29	46 (13.6)
Cough	27 (13.0)	12 (9.2)	.29	39 (11.6)
Malaise	12 (5.8)	6 (4.6)	.64	18 (5.3)
Local reaction	9 (4.3)	6 (4.6)	.90	15 (4.5)
Dizziness	9 (4.3)	3 (2.3)	.38	12 (3.6)
Mobility decreased	8 (3.9)	3 (2.3)	.54	11 (3.3)
Injection site swelling	5 (2.4)	5 (3.8)	.52	10 (3.0)

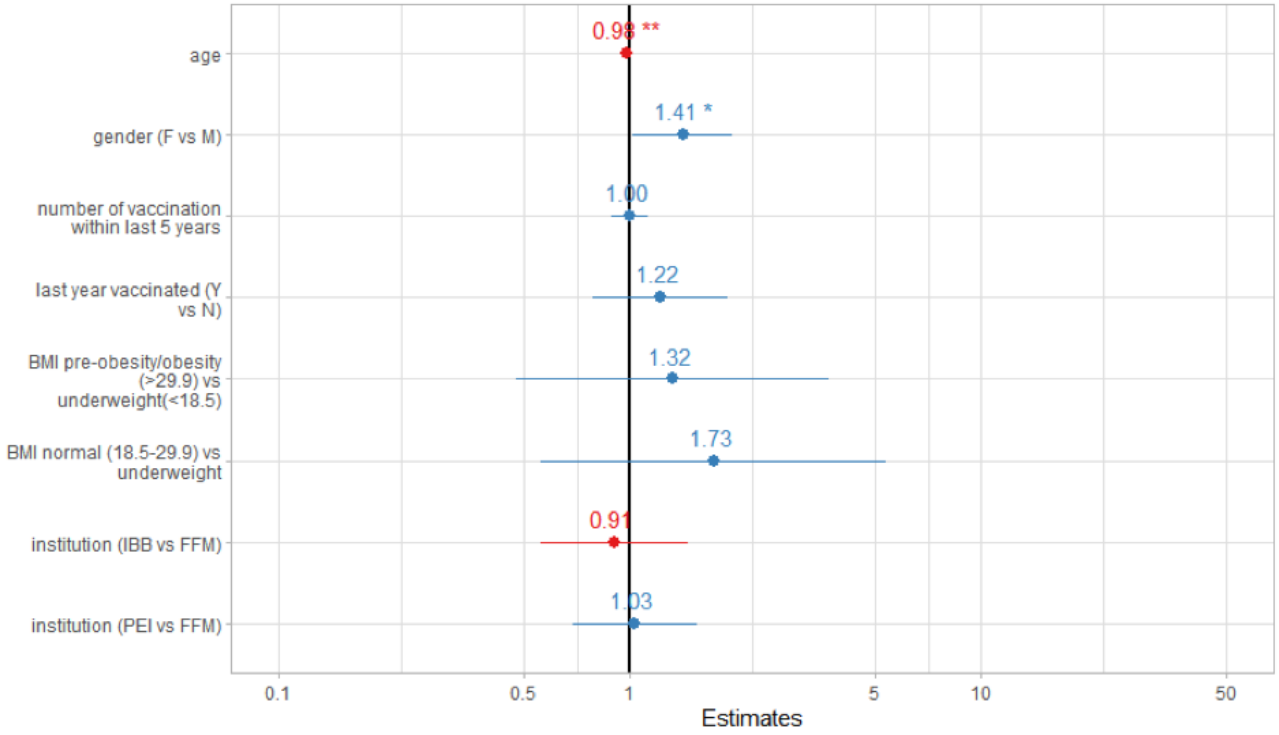
[a]AEFI: adverse event following immunization.

[b]Pearson chi-square test for cells n>5 and Fisher exact test for cells n≤5.

Factors Associated With AEFI Occurrence

For the outcome reported AEFI occurrence, the results of the Cox regression indicated a negative association with increasing age (hazard ratio 0.98; CI 0.97-0.99) and a positive association with female individuals (hazard ratio 1.41; CI 1.01-1.96; Figure 6).

Figure 6. Factors associated with adverse event following immunization occurrence versus nonoccurrence. F: female; M; male; Y: yes; N: no; IBB: Investitionsbank Berlin; FFM: University Hospital Frankfurt; PEI: Paul Ehrlich Institute.



Results of Additional Analysis

The hazard ratio for reporting an influenza-like illness at any time point within 3 months was 0.26 (CI 0.11-0.60) for persons who were vaccinated against influenza in the last year compared with those who were not vaccinated in the last year (Multimedia Appendix 1).

There were 6 users who responded to the usability questionnaire. The mean SUS score built out of 6 feedback questionnaire responders was 86.67 with no negative rating of the entertainment of the app, quality of information, or overall app suitability for reporting AEFIs.

Discussion

Principal Findings

Our study shows the feasibility of an app-based reporting of AEFIs and suggests an added value of a mobile app with near real-time features to report AEFIs for a period of 3 months. We found that adherence to app use was dependent on gender but independent of age and AEFI occurrence. On the other hand, determinants of a correct app entry were increasing age and higher education. The most observed AEFI types reported in our study were injection site pain, pain in extremity, and fatigue/asthenia.

Comparison With Prior Work

In previous studies, several methods for active adverse events reporting have been tested, including diary cards and telephone interviews and a mobile app [15,23]. Compared with a study using diary cards and telephone interviews [23], the dropout rate in our study was lower. The attrition rate in our study is consistent with the one reported by Wilson et al [15], who also used a mobile app as a reporting system for reporting AEFIs. The high dropout rate in the study using telephone interviews and diary cards could be related to a lack of active reminders. In our research and in Wilson et al [15], reminder notifications were implemented in the apps. This is of relevance especially for newly introduced vaccines for which the reporting period immediately after vaccination is often a crucial interval. For a 1-week interval, for example, our app showed an attrition rate of less than 2%. Common determinants of attrition in longitudinal studies are sex and age, with males and younger age being more likely to discontinue their participation [24,25]. However, our results showed that female sex was associated with attrition, whereas age had no influence. In our project, participants adapted to a new technology in order to stay adherent in a longitudinal study. The known factors associated with the nonadaption of new technologies are female sex and older age [26,27]. Therefore, it is not surprising that our results showed different determinants for attrition than in common longitudinal studies.

Vaccine-related information, such as the batch number, is often incorrectly entered when done manually [28]. Accordingly, one-third of our study participants reported they struggled with entering those details into the SafeVac app. For future studies or routine recording, the use of a mobile phone camera as barcode scanner could be an alternative to capture such information on vaccinations [29-31].

Older age of vaccinees, child age, and female sex are known triggers for AEFIs [32]. These particular age groups were not included in our study. This could explain why we found a decrease in reported AEFIs with increasing age. We cannot rule out to what extent AEFI occurrence is due to actual occurrence

or reporting behavior. However, this influences any kind of voluntary reporting.

In our study, we also asked participants to indicate previous influenza vaccinations. The uptake of the most recent previous vaccination was related to low reporting of an influenza-like illness as an AEFI within 3 months, whereas the overall number of influenza vaccination uptakes within the last 5 years seemed to have no effect.

The most observed types of AEFIs reported in our study (ie, injection site pain, pain in extremity, and fatigue/asthenia) correspond with other safety studies of influenza vaccinations [33]. One clinical trial, however, reported injection site pain, headaches, and myalgia more often (8% to 23% more often) than in our study [34]. The reasons for these differences are unclear. However, one possible reason could be that our study population might not be as healthy as the study population selected for the clinical trial.

When analyzing the kind of AEFI occurrence in relation to adherence level of app users/study participants (ie, fully and partially adherent users), we did not find a difference. Therefore, we conclude that no specific AEFI type was responsible for attrition. Furthermore, most of the attrition was toward the end of the study, leading to a similar likelihood of AEFI type occurrence in both groups.

Strengths and Limitations

The main strengths of our feasibility study are its multicentric study nature, long observation time, and use of a custom designed app for reporting AEFIs. However, the study did not aim to provide generalizable results. In our study, we could demonstrate that individuals are willing to report AEFIs for more than the usually implemented follow-up period of 2 weeks to 1 month for active influenza AEFI reporting [14,15]. With the help of additional studies, this could mean that a broader spectrum of AEFIs (ie, unexpected late-onset AEFIs) could be captured with an app. In addition, by using a longitudinal approach, we were able to generate determinants of adherence to app use and information on AEFIs occurring after an annual influenza vaccination. With another study design (eg, use of app in routine vaccination settings and for different vaccines), a more diverse and increased participant group (eg, elderly) could be captured. This would help to assess the causality of AEFIs, their frequencies, seriousness, and course. A general limitation in our findings is the fact that we cannot disentangle how and which determinants influence the reporting and factual occurrence of AEFIs.

Given that the mean SUS score of all respondents was more than 85/100, the app's usability seemed to be excellent according to the adjective scale rating [18]. However, we distributed the questionnaire to assess the usability and quality of the mobile app 1 year after recruitment due to difficulties in separating the feedback questionnaire from the national databank of PEI. The time delay between vaccination and questionnaire distribution could have led to a low response rate. In our study, we recruited participants face to face and used participation in a lottery as an incentive measure. Therefore, it is unclear if the same response and adherence rate can be expected outside of the study. Nevertheless, in order for participants to qualify for lottery participation, AEFI reports were required until 1 week after vaccination. As we do not see any immediate drop after that time point, we would exclude the incentive as the main reason behind the adherence to the app until the end of the study.

Additionally, our reported adverse event must be interpreted as such and not as an adverse reaction, meaning without any causality assessment. For that purpose, a comparison group with no vaccine would have been required.

Conclusions

We have shown that the use of a mobile app to report AEFIs for 3 months was feasible for more than 60% of participants, and the most reported adverse event after influenza vaccination were similar to those reported in clinical trials. Future studies could use the SafeVac concept for enhanced AEFI reporting, especially by including broader target groups (eg, elderly people and children) and by implementing the app in routine settings for various vaccines (eg, in general practitioners' offices). In addition, new vaccines like the one against COVID-19 can benefit from using an approach like SafeVac for safety reporting. In fact, SafeVac 2.0 was recently adapted to the COVID-19 vaccination and is used as the national reporting tool for AEFIs in an active surveillance study in Germany [35].

Acknowledgments

We would like to thank the occupational health physicians in charge at the 3 participating institutions, Investitionsbank Berlin, University Hospital Frankfurt, and PEI. We also thank all study participants for volunteering. We greatly appreciate the contributions of Monike Schlüter and Sabine Pape, who supported the field phase of the study and facilitated participant recruitment and logistics. The study was supported by the Federal Ministry of Health (Germany) and by intramural funds from the HZI (partly staff costs). The Federal Ministry of Health (Germany) and HZI had no role in data collection, analysis, or interpretation; study design; recruitment; or any aspect pertinent to the study. All authors had full access to the data in the study and accept responsibility to submit for publication.

Authors' Contributions

MTHN drafted the manuscript and conceived and performed the analysis plan with input from JJO and GK. JJO supervised the research and, together with GK, conceived the research questions and study design. DM and BKS provided guidance and expertise on pharmacovigilance and contributed to results interpretation. DM was responsible for tool development and data acquisition, and SG provided expertise on data analyses and verified the underlying data. All authors revised and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Factors associated with occurrence versus nonoccurrence of influenza-like illness.
[PNG File , 22 KB - mhealth_v9i5e26289_app1.png]

References

1. Immunization coverage. Geneva: World Health Organization URL: https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage [accessed 2021-04-06]
2. Vaccines. Geneva: World Health Organization URL: https://www.who.int/topics/vaccines/en/ [accessed 2021-04-06]
3. Ständige Impfkommission (STIKO) beim Robert Koch-Institut. [Ständige Impfkommission: Stellungnahme der Ständigen Impfkommission zu einer ku¨nftigen Impfung gegen COVID-19]. Epid Bull 2020;35:8-10 [FREE Full text] [doi: 10.25646/7090]
4. Pharmacovigilance Risk Assessment Committee (PRAC) Interim guidance on enhanced safety surveillance for seasonal influenza vaccines in the EU. 2014. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/04/WC500165492.pdf [accessed 2021-04-06]
5. Selecting viruses for the seasonal influenza vaccine. Atlanta: Centers for Disease Control and Prevention URL: https://www.cdc.gov/flu/prevent/vaccine-selection.htm [accessed 2021-04-06]
6. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 vaccines at pandemic speed. N Engl J Med 2020 May 21;382(21):1969-1973. [doi: 10.1056/NEJMp2005630] [Medline: 32227757]
7. Kahn R, Rid A, Smith PG, Eyal N, Lipsitch M. Choices in vaccine trial design in epidemics of emerging infections. PLoS Med 2018 Aug;15(8):e1002632 [FREE Full text] [doi: 10.1371/journal.pmed.1002632] [Medline: 30086139]
8. Isaacs D, Lawrence G, Boyd I, Ronaldson K, McEwen J. Reporting of adverse events following immunization in Australia. J Paediatr Child Health 2005 Apr;41(4):163-166. [doi: 10.1111/j.1440-1754.2005.00580.x] [Medline: 15813867]
9. Arzneimittelrisiken erfassen. Deutsche Apothekerzeitung. URL: https://www.deutsche-apotheker-zeitung.de/daz-az/2019/daz-37-2019/verdachtsfaelle-melden [accessed 2021-04-06]
10. Hasford J, Goettler M, Munter K, Müller-Oerlinghausen B. Physicians' knowledge and attitudes regarding the spontaneous reporting system for adverse drug reactions. J Clin Epidemiol 2002 Sep;55(9):945-950. [doi: 10.1016/s0895-4356(02)00450-x] [Medline: 12393084]
11. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
12. Waller PC. Measuring the frequency of adverse drug reactions. Br J Clin Pharmacol 1992 Mar;33(3):249-252 [FREE Full text] [doi: 10.1111/j.1365-2125.1992.tb04031.x] [Medline: 1576045]
13. The growing value of digital health. IQVIA. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2021-04-06]
14. Cashman P, Macartney K, Khandaker G, King C, Gold M, Durrheim DN. Participant-centred active surveillance of adverse events following immunisation: a narrative review. Int Health 2017 May 01;9(3):164-176 [FREE Full text] [doi: 10.1093/inthealth/ihx019] [Medline: 28582563]
15. Wilson K, Atkinson KM, Westeinde J, Bell C, Marty K, Fergusson D, et al. An evaluation of the feasibility and usability of a proof of concept mobile app for adverse event reporting post influenza vaccination. Hum Vaccin Immunother 2016 Jul 02;12(7):1738-1748 [FREE Full text] [doi: 10.1080/21645515.2016.1152434] [Medline: 26905396]
16. Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. J Thorac Dis 2015 Nov;7(11):E537-E540 [FREE Full text] [doi: 10.3978/j.issn.2072-1439.2015.10.63] [Medline: 26716051]
17. Lewis JR. The System Usability Scale: past, present, and future. Int J Human Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
18. Bangor A, Staff T, Kortum P, Miller J, Staff T. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
19. Nguyen MTH, Ott JJ, Caputo M, Keller-Stanislawski B, Klett-Tammen CJ, Linnig S, et al. User preferences for a mobile application to report adverse events following vaccination. Pharmazie 2020 Jan 02;75(1):27-31. [doi: 10.1691/ph.2020.9734] [Medline: 32033630]
20. Body Mass Index—BMI. Geneva: World Health Organization URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [accessed 2021-04-06]
21. Brooke J. SUS: a quick and dirty usability scale. Usability Eval Ind 1996;189(194):4-7. [doi: 10.1201/9781498710411-35]
22. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]

23. Wu J, Xu F, Lu L, Lu M, Miao L, Gao T, et al. Safety and effectiveness of a 2009 H1N1 vaccine in Beijing. N Engl J Med 2010 Dec 16;363(25):2416-2423. [doi: 10.1056/NEJMoa1006736] [Medline: 21158658]
24. Fischer EH, Dornelas EA, Goethe JW. Characteristics of people lost to attrition in psychiatric follow-up studies. J Nerv Ment Dis 2001 Jan;189(1):49-55. [doi: 10.1097/00005053-200101000-00009] [Medline: 11206665]
25. Bellón JA, de Dios Luna J, Moreno B, Montón-Franco C, GildeGómez-Barragán MJ, Sánchez-Celaya M, et al. Psychosocial and sociodemographic predictors of attrition in a longitudinal study of major depression in primary care: the predictD-Spain study. J Epidemiol Community Health 2010 Oct;64(10):874-884. [doi: 10.1136/jech.2008.085290] [Medline: 19759057]
26. Morris M, Venkatesh V, Ackerman P. Gender and age differences in employee decisions about new technology: an extension to the theory of planned behavior. IEEE Trans Eng Manag 2005 Feb;52(1):69-84. [doi: 10.1109/tem.2004.839967]
27. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
28. Glöckner S. Assessment of data quality in the vaccination registry of the German National Cohort (GNC). URL: https://www.dgepi2018.de/wp-content/uploads/2018/09/Entwurf_Abstractbook_DGepi2018_SS_24092018.pdf [accessed 2021-04-06]
29. Marques J, Ribeiro-Vaz I, Pereira AC, Polónia J. A survey of spontaneous reporting of adverse drug reactions in 10 years of activity in a pharmacovigilance centre in Portugal. Int J Pharm Pract 2014 Aug;22(4):275-282. [doi: 10.1111/ijpp.12078] [Medline: 24188533]
30. Pereira JA, Quach S, Hamid JS, Heidebrecht CL, Quan SD, Nassif J, et al. Exploring the feasibility of integrating barcode scanning technology into vaccine inventory recording in seasonal influenza vaccination clinics. Vaccine 2012 Jan 17;30(4):794-802. [doi: 10.1016/j.vaccine.2011.11.043] [Medline: 22119585]
31. Pereira JA, Quach S, Hamid JS, Quan SD, Diniz AJ, Van ER, et al. The integration of barcode scanning technology into Canadian public health immunization settings. Vaccine 2014 May 13;32(23):2748-2755 [FREE Full text] [doi: 10.1016/j.vaccine.2013.11.015] [Medline: 24252700]
32. Alomar MJ. Factors affecting the development of adverse drug reactions (Review article). Saudi Pharm J 2014 Apr;22(2):83-94 [FREE Full text] [doi: 10.1016/j.jsps.2013.02.003] [Medline: 24648818]
33. de Graaf R, van Dorsselaer S, Tuithof M, ten Have M. Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands Mental Health Survey and Incidence Study-2. Compr Psychiatry 2013 Nov;54(8):1131-1139. [doi: 10.1016/j.comppsych.2013.05.012] [Medline: 23810078]
34. Gresset-Bourgeois V, Leventhal PS, Pepin S, Hollingsworth R, Kazek-Duret M, De Bruijn I, et al. Quadrivalent inactivated influenza vaccine (VaxigripTetra™). Expert Rev Vaccines 2018 Jan;17(1):1-11. [doi: 10.1080/14760584.2018.1407650] [Medline: 29157068]
35. Paul-Ehrlich-Institut. SafeVac 2.0 Smartphone app to survey tolerability of COVID-19 vaccines. URL: https://www.pei.de/EN/newsroom/hp-news/2020/201222-safevac-app-smartphone-survey-tolerability-covid-19-vaccines.html [accessed 2021-04-06]

Abbreviations

AEFI: adverse event following immunization
OR: odds ratio
PEI: Paul Ehrlich Institute
SUS: System Usability Scale

Edited by L Buis; submitted 10.12.20; peer-reviewed by G Akpan, K Wilson, D Pförringer; comments to author 25.01.21; revised version received 05.02.21; accepted 12.03.21; published 07.05.21.

Please cite as:
Nguyen MTH, Krause G, Keller-Stanislawski B, Glöckner S, Mentzer D, Ott JJ
Postmarketing Safety Monitoring After Influenza Vaccination Using a Mobile Health App: Prospective Longitudinal Feasibility Study
JMIR Mhealth Uhealth 2021;9(5):e26289
URL: https://mhealth.jmir.org/2021/5/e26289
doi:10.2196/26289
PMID:33960950

Original Paper

Feasibility and Acceptability of an Asthma App to Monitor Medication Adherence: Mixed Methods Study

Cristina Jácome[1,2], PhD; Rute Almeida[1,2], PhD; Ana Margarida Pereira[1,2,3], MD; Rita Amaral[1,2,4,5], PhD; Sandra Mendes[1], MSc; Magna Alves-Correia[3], MD; Carmen Vidal[6], MD, PhD; Sara López Freire[6], MD; Paula Méndez Brea[6], MD; Luís Araújo[3], MD; Mariana Couto[3], MD, PhD; Darío Antolín-Amérigo[7], MD, PhD; Belén de la Hoz Caballer[7], MD, PhD; Alicia Barra Castro[7], MD; David Gonzalez-De-Olano[7], MD, PhD; Ana Todo Bom[8], MD, PhD; João Azevedo[9], MD; Paula Leiria Pinto[10], MD, MSc; Nicole Pinto[10], MD; Ana Castro Neves[10], MD; Ana Palhinha[10], MD; Filipa Todo Bom[11], MD; Alberto Costa[12], MD; Cláudia Chaves Loureiro[13], MD, PhD; Lilia Maia Santos[14], MD; Ana Arrobas[13], MD; Margarida Valério[13], MD; João Cardoso[15], MD, PhD; Madalena Emiliano[15], MD; Rita Gerardo[15], MD; José Carlos Cidrais Rodrigues[16], MD; Georgeta Oliveira[16], MD; Joana Carvalho[16], MD; Ana Mendes[17], MD; Carlos Lozoya[18], MD, PhD; Natacha Santos[19], MD; Fernando Menezes[20], MD; Ricardo Gomes[20], MD; Rita Câmara[21], MD; Rodrigo Rodrigues Alves[22], MD; Ana Sofia Moreira[22], MD; Diana Bordalo[1,23], MD; Carlos Alves[24], MD, MSc; José Alberto Ferreira[25], MD; Cristina Lopes[26,27], MD, PhD; Diana Silva[28], MD, PhD; Maria João Vasconcelos[28], MD; Maria Fernanda Teixeira[29], MD; Manuel Ferreira-Magalhães[1,2,29], MD, PhD; Luís Taborda-Barata[30,31], MD, PhD; Maria José Cálix[32], MD; Adelaide Alves[33], MD; João Almeida Fonseca[1,2,3,34], MD, PhD

[1]Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal

[2]Department of Community Medicine, Faculty of Medicine, Information and Health Decision Sciences, University of Porto, Porto, Portugal

[3]Allergy Unit, Instituto and Hospital CUF, Porto, Portugal

[4]Department of Cardiovascular and Respiratory Sciences, Porto Health School, Porto, Portugal

[5]Department of Women's and Children's Health, Paediatric Research, Uppsala University, Uppsala, Sweden

[6]Servicio de Alergia, Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain

[7]Servicio de Alergia, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

[8]Serviço de Imunoalergologia, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

[9]Imunoalergologia, Centro Hospitalar de Leiria, Leiria, Portugal

[10]Serviço de Imunoalergologia, Hospital de Dona Estefânia, Centro Hospitalar Universitário de Lisboa Central, Lisboa, Portugal

[11]Serviço de Pneumologia, Hospital Beatriz Ângelo, Loures, Portugal

[12]Serviço de Pediatria, Hospital da Senhora da Oliveira, Guimarães, Portugal

[13]Serviço Pneumologia, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

[14]Serviço de Pneumologia, Hospital Distrital da Figueira da Foz, Figueira da Foz, Portugal

[15]Serviço de Pneumologia, Hospital Santa Marta, Centro Hospitalar Universitário de Lisboa Central, Lisboa, Portugal

[16]Serviço de Pediatria, Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal

[17]Serviço de Imunoalergologia, Hospital de Santa Maria, Centro Hospitalar Universitário Lisboa Norte, Lisboa, Portugal

[18]Serviço de Imunoalergologia, Hospital Amato Lusitano, Unidade Local de Saúde de Castelo Branco, Castelo Branco, Portugal

[19]Serviço de Imunoalergologia, Centro Hospitalar Universitário do Algarve, Portimão, Portugal

[20]Serviço de Pneumologia, Hospital Garcia de Orta, Almada, Portugal

[21]Serviço de Imunoalergologia, Serviço de Saúde da Região Autónoma da Madeira, Funchal, Portugal

[22]Serviço de Imunoalergologia, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal

[23]Serviço de Pediatria, Unidade Hospitalar de Famalicão, Centro Hospitalar do Médio Ave, Vila Nova de Famalicão, Portugal

[24]Serviço de Pneumologia, Hospital Nossa Senhora do Rosário, Centro Hospitalar Barreiro Montijo, Barreiro, Portugal

[25]Serviço de Imunoalergologia, Unidade I, Centro Hospitalar Vila Nova de Gaia/Espinho, Vila Nova de Gaia, Portugal

[26]Unidade de Imunoalergologia, Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal

[27]Imunologia Básica e Clínica, Faculdade de Medicina, Universidade do Porto, Porto, Portugal

[28]Serviço de Imunoalergologia, Centro Hospitalar Universitário de São João, Porto, Portugal

[29]Serviço de Pediatria, Centro Materno Infantil do Norte, Centro Hospitalar Universitário do Porto, Porto, Portugal

[30]Department of Allergy & Clinical Immunology, Cova da Beira University Hospital Centre, Covilhã, Portugal

[31]Environment & Health Study Group, Faculty of Health Sciences, Health Sciences Research Centre, University of Beira Interior, Covilhã, Portugal

[32]Serviço de Pediatria, Hospital de São Teotónio, Centro Hospitalar Tondela–Viseu, Viseu, Portugal
[33]Serviço de Pneumologia, Unidade I, Centro Hospitalar Vila Nova de Gaia/Espinho, Vila Nova de Gaia, Portugal
[34]Medicina, Educação, Investigação, Desenvolvimento e Avaliação, Porto, Portugal

Corresponding Author:
Cristina Jácome, PhD
Center for Health Technology and Services Research
Faculty of Medicine
University of Porto
Rua Dr Plácido da Costa
Porto, 4200-450
Portugal
Phone: 351 225 513 622
Fax: 351 225 513 668
Email: cristinajacome.ft@gmail.com

Abstract

Background: Poor medication adherence is a major challenge in asthma, and objective assessment of inhaler adherence is needed. The InspirerMundi app aims to monitor adherence while providing a positive experience through gamification and social support.

Objective: This study aimed to evaluate the feasibility and acceptability of the InspirerMundi app to monitor medication adherence in adolescents and adults with persistent asthma (treated with daily inhaled medication).

Methods: A 1-month mixed method multicenter observational study was conducted in 26 secondary care centers from Portugal and Spain. During an initial face-to-face visit, physicians reported patients' asthma therapeutic plan in a structured questionnaire. During the visits, patients were invited to use the app daily to register their asthma medication intakes. A scheduled intake was considered taken when patients registered the intake (inhaler, blister, or other drug formulation) by using the image-based medication detection tool. At 1 month, patients were interviewed by phone, and app satisfaction was assessed on a 1 (low) to 5 (high) scale. Patients were also asked to point out the most and least preferred app features and make suggestions for future app improvements.

Results: A total of 107 patients (median 27 [P25-P75 14-40] years) were invited, 92.5% (99/107) installed the app, and 73.8% (79/107) completed the 1-month interview. Patients interacted with the app a median of 9 (P25-P75 1-24) days. At least one medication was registered in the app by 78% (77/99) of patients. A total of 53% (52/99) of participants registered all prescribed inhalers, and 34% (34/99) registered the complete asthma therapeutic plan. Median medication adherence was 75% (P25-P75 25%-90%) for inhalers and 82% (P25-P75 50%-94%) for other drug formulations. Patients were globally satisfied with the app, with 75% (59/79) scoring ≥4,; adherence monitoring, symptom monitoring, and gamification features being the most highly scored components; and the medication detection tool among the lowest scored. A total of 53% (42/79) of the patients stated that the app had motivated them to improve adherence to inhaled medication and 77% (61/79) would recommend the app to other patients. Patient feedback was reflected in 4 major themes: medication-related features (67/79, 85%), gamification and social network (33/79, 42%), symptom monitoring and physician communication (21/79, 27%), and other aspects (16/79, 20%).

Conclusions: The InspirerMundi app was feasible and acceptable to monitor medication adherence in patients with asthma. Based on patient feedback and to increase the registering of medications, the therapeutic plan registration and medication detection tool were redesigned. Our results highlight the importance of patient participation to produce a patient-centered and engaging mHealth asthma app.

(JMIR Mhealth Uhealth 2021;9(5):e26442) doi:10.2196/26442

KEYWORDS

mHealth; smartphone; technology assessment; medication adherence; self-management; gamification; patient participation

Background

Inhaled preventive medications are the cornerstone of effective asthma treatment, reducing symptoms and exacerbations [1]. Yet in the last three decades, rates of adherence among patients with asthma remained unchanged, ranging from 19% to 64% [2]. Poor medication adherence is thus a major challenge in asthma [2,3] and is a leading driver of poor health outcomes, including emergency department visits, hospitalizations, and increased health care costs [4]. Based on this significant burden, medication adherence is a top priority for disease management [1] and policy agendas [5,6].

Research efforts to tackle medication nonadherence have not led to the much needed improvement [7]. Behavioral and

educational interventions had modest effects, possibly because most propose one-size-fits-all solutions for distinct nonadherence phenotypes (erratic, unwitting, intelligent) [8]. These interventions also failed to consider facilitators for real-world implementation [8]. The objective measurement of inhaler adherence is another major problem [9]. Electronic monitoring devices are available but their dissemination in real-world practice is difficult and costly [10].

Mobile health (mHealth) technologies may be promising to support medication adherence as they are easily integrated into patients' everyday lives, especially if using only smartphone embedded sensors [11]. mHealth allows the combination of computation, communication, and interactive display that makes it possible to use gamification and peer support tools with the potential to inspire behavior changes and offer treatment adherence quantification [11,12]. Moreover, patients with asthma have increasingly high smartphone ownership levels and are interested in using disease-specific apps [13,14].

The InspirerMundi app development was grounded in previous research and cooperation with patients with asthma and physicians [15]. The app aims to transform adherence to inhalers into a positive experience through gamification and social support while allowing for verified inhaler adherence monitoring. The concept and gameplay of the InspirerMundi app derive from promoting user social interaction in a quest to help other users increase inhaler adherence (while being an example of good adherence). The user's goal is to become an Inspirer to increase the sphere of positive influence by coaching an expanding network of Warriors (as Warriors progress and keep good medication adherence, they can become Inspirers themselves). The most innovative app technology is the medication detection tool based on advanced processing of inhaler images captured with the smartphone camera [16].

The app development process was iterative with real-life tests and improvements based on user feedback. The first version was previously tested in a usability study, and improvements were made [17]. It is believed that the InspirerMundi app is innovative in measuring and improving medication adherence in patients with asthma. Before studying its effectiveness, however, there is a need for feasibility studies to monitor its use and evaluate its acceptability. Thus, this study aimed to evaluate the feasibility and acceptability of the InspirerMundi app to monitor medication adherence in adolescents and adults with asthma.

Methods

Study Design

A multicenter observational study with one initial face-to-face visit and a telephone interview at 1 month was conducted [18]. During the study, an additional telephone interview at 1 week for additional data collection and 3 text messages (at 2, 14, and 21 days) to promote app engagement were added. A convenience sample of adolescents and adults with persistent asthma was recruited between November 2017 and April 2019 at 26 allergy, pulmonology, and pediatric secondary care centers in Portugal (North, Centre, Lisbon, Algarve, Azores, and Madeira regions) and Spain (Galicia and Madrid regions). Centers were asked to recruit at least 2 patients as a minimum of 20 adolescents and 20 adults were targeted [19,20]. The study protocol was approved by the ethics committees of all participating centers. The study was conducted in accordance with the ethical standards established in the Declaration of Helsinki. Eligible patients were approached by physicians during medical visits and invited to participate in a study on the use of an app to register their asthma medication daily intake. Written informed consent was obtained before enrollment in the study. Adult patients signed a consent form; adolescents signed an assent form and a parental consent form was also obtained. The study is reported according to STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [21].

Participants

Patients were included if they (1) had a previous medical diagnosis of persistent asthma, (2) were at least 13 years old (13 to 17 years for adolescents; 18 years and older for adults), (3) were able to use mobile apps and had access to a mobile device with internet access, and (4) had an active prescription for a daily inhaled controller medication for asthma. All inhaled controller treatments were allowed, and there was no change in any prescribed medication as a result of the participation in this study. Patients were excluded if they had a diagnosis of a chronic lung disease other than asthma or a diagnosis of another significant chronic condition with possible interference with the study aims.

InspirerMundi App

The InspirerMundi app is focused on supporting patient medication management and promoting treatment adherence. The InspirerMundi app is grounded on the Fogg behavior model, which states that a behavior is performed when 3 elements converge in a given moment: motivation, ability, and trigger (Table 1) [22]. The app includes registration of the therapeutic plan, which activates notifications when a medication is due (trigger) and allows recording of performed intakes through a medication detection tool that uses the smartphone camera. The medication detection tool is based on advanced image processing techniques and identifies the inhaler through template matching [15]. At the time of the study, the tool could detect 6 inhaler devices containing numeric dose counters (Diskus/Accuhaler [GlaxoSmithKline PLC], Ellipta [GlaxoSmithKline PLC], Flutiform pMDI [Napp Pharmaceuticals LTD], Novolizer/Genuair [AstraZeneca], Spiromax [Teva Pharmaceutical Industries LTD], Turbohaler [AstraZeneca]). The medication detection tool can also be applied to other types of inhalers or other drug formulations (eg, blister or other recipients) but the template matching feature cannot. In the app version tested, a scheduled intake was considered taken when patient presented the medication (inhaler, blister, or other drug formulation) to the detection tool (10 seconds). Adherence statistics are provided to the user in the form of circular progress graphs (motivator: self-monitoring; ability: simplifies adherence measurement). Besides planned medication, the user may also record relief medication intake events that are not considered for treatment adherence assessment. Physician involvement is also supported by providing the user with the possibility of

sharing the registered therapeutic plan and respective adherence data. Medication management is supported by a timeline that reflects the registered therapeutic plan. The timeline also includes events related to gamification and symptom monitoring. Patients are invited to answer 3 types of questionnaires related to their symptoms and asthma control: daily questionnaire (daily after 6 PM), weekly questionnaire (once per week), and Control of Allergic Rhinitis and Asthma Test (CARAT, by default once per month but can be personalized by the patient to weekly or every 2 weeks) [23] (motivator: self-monitoring, personalization; ability: simplifies assessment of asthma control). CARAT total score is calculated by summing the 10 questions resulting in a range of 0 to 30 points. A score >24 indicates good disease control [24], and this classification is presented to the patient (motivator: anticipation/hope of a health improvement). This questionnaire is available in multiple languages, and its psychometric properties are established [25].

The concept and gamification dynamics are twofold: achievement challenges and social interaction (motivator: belonging, social acceptance, app as a social actor). Gameplay and mechanics included in the app derive from achievement points and badges (triggers) and by promoting social interaction based on a story of evolving from a Warrior (beginner player, level 1) to an Inspirer (advanced player, an example of good adherence, available from level 10 onward), who can motivate Warriors to improve adherence (motivators: cooperation, competition, recognition). The main components that support the game are points earned when users take certain actions (motivator: sensation of pleasure, positive reward) such as registering a new medication, registering scheduled medication intakes at the right time, answering symptom questionnaires, or getting a positive assessment from other users in their network. Registering events of relief medication does not generate additional points. When registering a relief medication intake, patients are invited to rate the severity of symptoms experienced on a visual analog scale (VAS; 0 to 100) and report any use of health care. Another encouragement for user interaction is the attribution of virtual badges (12 available; motivator: sensation of pleasure, positive reward). Whenever the user reaches a certain goal (eg, Role Model badge when the user links to the first Warrior) or does something special (eg, Big Influencer badge when the user reaches 5 Warriors), they will be rewarded and get a virtual badge for their actions. The app, developed in Portuguese, English, and Spanish, is available in the App Store (version 1.1) and Google Play (versions 1.1.x); app is only available for download in Portugal and Spain. Screenshots of the app version tested (version 1.1) are presented in Figure 1, and videos are available [26-29].

Figure 1. InspirerMundi app screenshots (version 1.1).

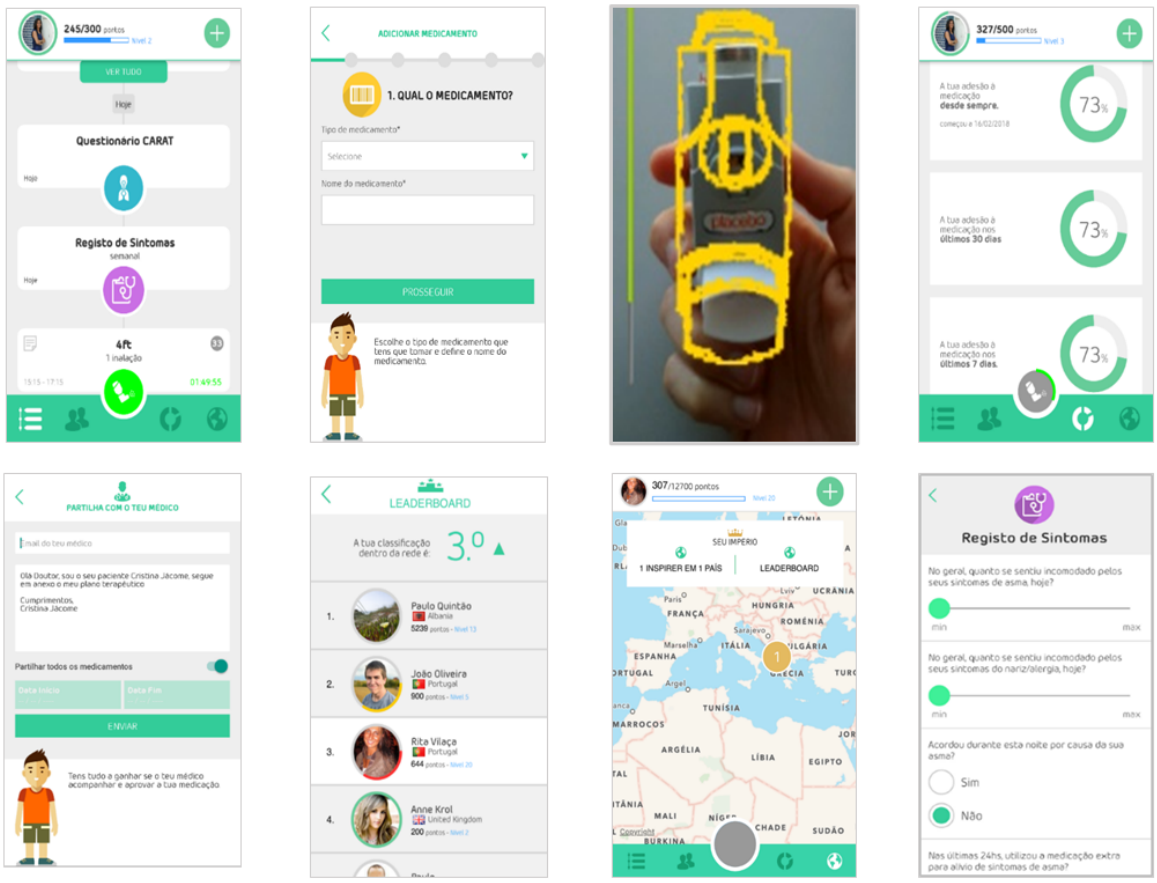


Table 1. Elements of the Fogg behavior model in the InspirerMundi app.

Elements and description of the Fogg behavior model	App feature	App component
Trigger		
Facilitator of adherence	Signal: notification appears when a medication intake is due	Medication management
Facilitator of medication intake registration	Signal: intake icon flashes when a medication intake is due	Medication management
Facilitator of asthma control assessment	Signal: questionnaire icon flashes when a response is due	CARAT[a], daily questionnaire, and weekly questionnaire
Facilitator of adherence	Signal: buzzes when an Inspirer sends a medication reminder to a Warrior	Medication management, gamification
Facilitator of adherence	Signal: icon of message appears (exchange between Inspirer and Warrior)	Gamification
Ability		
Simplifies adherence measurement	Adherence statistics: circular progress graph	Medication management
Simplifies assessment of asthma control	Asthma control classification	CARAT
Motivator		
Self-monitoring	Adherence statistics: circular progress graph	Medication management
Self-monitoring, personalization	Personalized frequency	CARAT
Anticipation/hope of a health improvement	Asthma control classification	CARAT
Belonging, social acceptance, app as a social actor	Social interaction	Gamification
Cooperation, competition, recognition	Inspirer and Warrior players	Gamification
Sensation of pleasure, positive reward	Points	Gamification
Sensation of pleasure, positive reward	Virtual badges	Gamification

[a]CARAT: Control of Allergic Rhinitis and Asthma Test.

Data Collection

During the initial face-to-face visit, physician recorded if it was a first or follow-up appointment; patient's asthma control according to the Global Initiative for Asthma (GINA) [1]; number of exacerbations in the previous year (defined as episodes of progressive increase in shortness of breath, cough, wheezing, and/or chest tightness requiring a change in maintenance therapy [30]); and use of health care resources in the previous year (ie, number of unscheduled medical visits to primary care, secondary care, or emergency department and number of hospital admissions). Physician also recorded patient's current asthma treatment, including inhaled and oral medication, allergen immunotherapy, and biologic therapy. At the visit, patient answered written questionnaires on demographic data (age, gender, BMI, smoking habits), adherence to inhaled controller asthma medication during the previous week, and perception of the correctness of inhaler technique (using a 100 mm VAS), asthma control during the previous 4 weeks (using CARAT), perception of their overall health (using the EuroQol 5-Dimension [EQ-5D] VAS [31]), and previous use of health and fitness apps.

At 1 week and 1 month, patients were asked in a telephone interview about their adherence to inhaled controller medication during the previous week (on a 0 to 100 scale) and their asthma control with CARAT. At 1 month, they were asked to rate the app usability using 4 items of the System Usability Scale (ease of use, function integration, confidence in using, would use frequently) [32]. Patients were also asked about their general satisfaction with the app and app components (gamification, peer support, symptom monitoring, medication detection tool, adherence monitoring), their willingness to recommend the app to others, and the impact of the app on the motivation to adhere to the inhaler. All these satisfaction questions were assessed using a 1 (low) to 5 (high) scale. Patients were asked open questions on the most and least preferred app features and asked to make suggestions for app improvements. These telephone interviews could also provide support regarding any technical issues with the app or study-related doubts. Nonscheduled telephone contacts could also occur, as patients were provided with email and telephone contacts to use in case of any app technical issue; this was rarely used. Interviews were performed by a central team of trained health professionals and lasted around 15 minutes. An interview guide was used to standardize data collection across all patients.

Data Analysis

Descriptive statistics were used to characterize the sample and app use. The normality of each variable was investigated with Kolmogorov-Smirnov tests and visual analysis of histograms. The app use rate was taken as the ratio between the number of days with app use and the period of use (31-day period). Adherence to medication measured by the app was calculated using 2 methods: considering the medication taken only on days

with app use and considering the medication taken regardless of app use (ie, considering all medication scheduled for the 31-day period). Comparison of the characteristics of patients installing the app versus those not installing, using the app for more than 1 day versus single-day use, completing the study versus dropouts, and adolescents versus adults were performed using independent samples *t* tests, Mann-Whitney *U* tests for continuous variables, and chi-square tests for categorical variables. A Fisher test was used to compare adherence to inhalers measured by the app with patient self-reports (at the initial visit, 1 week, and 1 month). The Wilcoxon test and McNemar test were used to analyze differences in CARAT total score and CARAT control classification, respectively. Statistical analyses were performed using SPSS (version 26.0, IBM Corp) and plots were created using GraphPad Prism (version 6.0, GraphPad Software). The level of significance was set at .05. A thematic qualitative analysis of patient feedback (ie, opinions, suggestions) about the app collected during the 1-month telephone interview was also conducted. Notes from the interviews were read, and emerging themes were grouped. Description of patient feedback in major themes is supported with translated representative statements.

Results

Participants

A total of 107 patients were included: 77 from 24 Portuguese centers and 30 from 2 Spanish centers. Patients included mostly adults (77/107; 72.0%) and females (63/107; 58.9%). Most were on inhaled corticosteroid/long-acting beta-agonist combination therapy (90/107; 84.1%) and used only one inhaler (67/107; 62.6%). Based on GINA classification, almost half (52/107; 48.6%) of the participants did not have their asthma well controlled. Characteristics of the participants are shown in Table 2.

Table 2. Participant baseline characteristics (n=107).

Characteristic	Total (n=107)	Did not install app (n=8)	App used 1 day (n=25)	App used >1 day (n=74)
Age (years), median (P25-P75[a])	27 (17-40)	40 (28-46)	22 (16-40)	26.5 (17-39)
Adults, n (%)	77 (72.0)	8 (100)	15 (60)	54 (73)
Female, n (%)	63 (58.9)	6 (75)	10 (40)	47 (64)
BMI (kg/m^2), median (P25-P75)	23.1 (21-27)	26.2 (22-27)	24 (22-29)	22.8 (21-26)
Smoking status, n (%)				
Never smoker	79 (73.8)	5 (63)	17 (68)	57 (77)
Ex-smoker	19 (17.8)	3 (37)	5 (20)	11 (15)
Current smoker	9 (8.4)	0	3 (12)	6 (8)
Inhaled medication, n (%)				
ICS[b]/LABA[c]	90 (84.1)	6 (75)	23 (92)	61 (82)
SABA[d]	26 (24.3)	3 (38)	9 (36)	14 (19)
ICS	13 (12.1)	1 (13)	2 (8)	10 (14)
LAMA[e]	10 (9.3)	0	1 (4)	9 (12)
Single inhaler, n (%)	67 (62.6)	4 (50)	15 (60)	48 (65)
VAS[f] self-reported inhaler adherence (0-100)[g]	80 (60-90)	83 (59-91)	79 (53-90)	85 (65-90)
High (81-100), n (%)	51 (47.7)	4 (50)	7 (28)	40 (54)
Medium (51-80), n (%)	37 (34.6)	3 (38)	12 (48)	22 (30)
Low (0-50), n (%)	15 (14.0)	1 (12)	4 (16)	10 (14)
VAS correctness of inhaler technique (0-100)	99 (90-100)	92.5 (89-99)	90 (80-100)	99.5 (90-100)
Oral medication, antileukotriene, n (%)	42 (39.3)	0	8 (32)	34 (46)
Allergen immunotherapy, n (%)	19 (17.8)	2 (25)	5 (20)	12 (16)
Biologic therapy, n (%)	8 (7.5)	0	1 (4)	7 (9)
GINA[h] assessment symptom control[i], n (%)				
Well controlled	53 (49.5)	3 (38)	11 (44)	39 (53)
Partly controlled/uncontrolled	52 (48.6)	4 (50)	13 (52)	35 (47)
CARAT[j] total score, mean (SD)	21 (17-23)	14.5 (9-27)	20 (16-22)	21 (17-24)
≥1 exacerbation past year, n (%)	52 (48.6)	6 (75)	16 (64)	30 (41)
≥1 unscheduled medical visit past year, n (%)	27 (25.2)	2 (25)	10 (40)	15 (20)
EQ-5D[k] VAS, mean (SD)	80 (70-90)	80 (58-89)	75 (65-85)	82 (70-90)
First medical visit, n (%)	23 (21.5)	4 (50)	8 (32)	11 (15)
Previous use of health and fitness apps, n (%)	58 (54.2)	5 (63)	12 (50)	41 (55)

[a]P25-P75: percentile 25 to percentile 75.

[b]ICS: inhaled corticosteroid.

[c]LABA: long-acting beta-agonist.

[d]SABA: short-acting beta-agonist.

[e]LAMA: long-acting muscarinic receptor antagonist.

[f]VAS: visual analog scale.

[g]4 missing values.

[h]GINA: Global Initiative for Asthma.

[i]2 missing values.

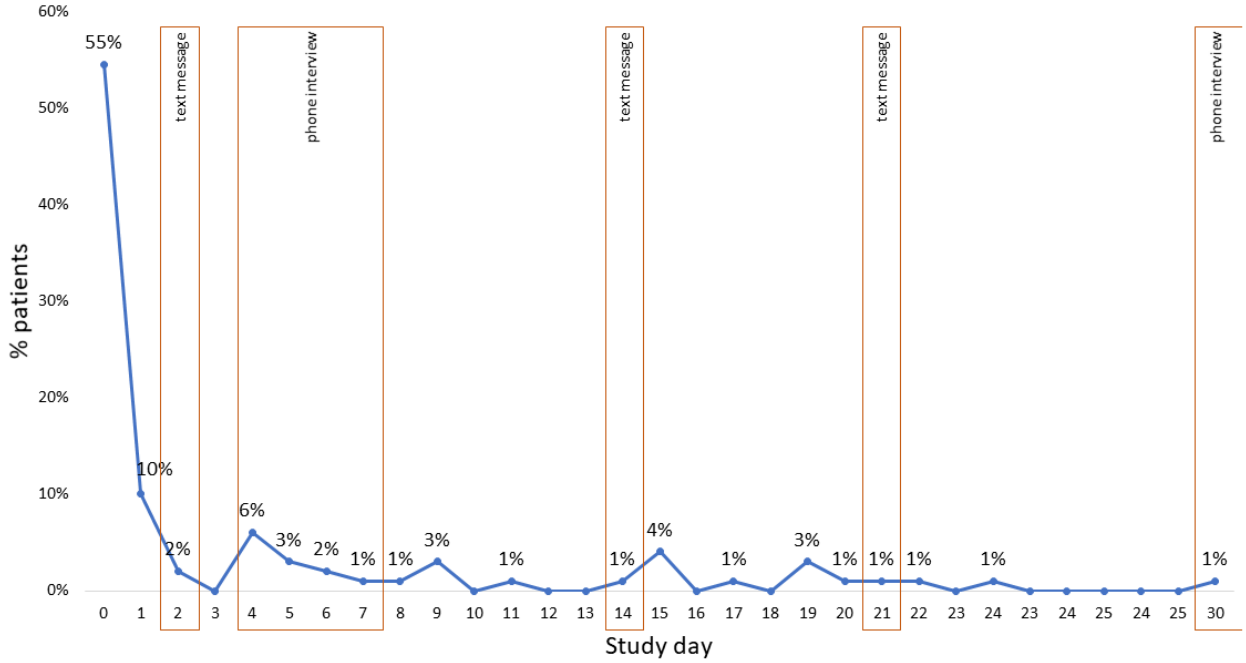
[j]CARAT-T: Control of Allergic Rhinitis and Asthma Test total score.

[k]EQ-5D: EuroQol 5-Dimension test.

App Use

In total, 92.5% (99/107) of patients installed the app. The 8 patients who did not install the app were more frequently in a first medical appointment with that physician (*P*=.02) and tended to be older (*P*=.05) than those who installed the app. No other significant differences were observed between the 2 groups. More than half (64/99; 65%) of patients installed the app on the recruitment day or the following day (median 0 [P25-P75 0-5]; Figure 2).

Figure 2. Percentage of users installing the app per study day.



Patients interacted with the app a median of 9 (P25-P75 1-24) days: 25% (25/99) only 1 day and 10% (10/99) all 31 days (Figure 3). A median use rate of 29% (P25-P75 3%-77%) was observed. Patients interacting only 1 day were more frequently male (*P*=.04), had a worse perception of their health status (*P*=.045), and had more exacerbations in the previous year (*P*=.04) in comparison with patients interacting more than 1 day. Although not reaching statistical significance, these patients had a higher need for unscheduled medical visits in the previous year (*P*=.05), more frequent short-acting beta-agonist prescribed in their therapeutic plan (*P*=.09, physician-reported), and more frequent depression symptoms (*P*=.09).

A total of 78% (77/99) of patients scheduled medication in the app (median 2 [P25-P75 1-3] medications) for 25 (P25-P75 17-31) days: 74% (73/99) registered at least 1 inhaler and 42% (42/99) at least another medication (eg, pills, nasal spray). The asthma therapeutic plan reported by the physician included a median of 2 (P25-P75 1.8-2) medications. A total of 53% (52/99) of participants registered all prescribed inhalers and 34% (34/99) registered the complete asthma therapeutic plan (inhaler and other drug formulations). Two (2%) patients also used the app to register medications not related to asthma.

Median inhaler adherence assessed through the app was 81% (P25-P75 58%-92%) when considering only days with app use and 62% (P25-P75 23%-81%) when considering all scheduled inhalations for the 31-day period (*P*<.001). Higher estimates were observed when patients self-reported adherence at initial visit (85 [P25-P75 63-90]), 1-week (95 [P25-P75 80-100]) and 1-month (100 [P25-P75 90-100]) interviews (*P*<.001; all pairwise comparisons *P*<.009 with exception of 1-week and 1-month interview that were not different). Median adherence for other drug formulations was 83% (P25-P75 50%-94%) when considering only days with app use and 73% (P25-P75 26%-94%) when considering all scheduled inhalations for the 31-day period (*P*<.001).

A total of 17% (17/99) registered the use of relief medication for at least 1 day (median 1 [P25-P75 1-2]): 13% (13/99) used inhalers, 4% (4/99) other drug formulations (pills, nasal spray), and 1% (1/99) inhaler plus others. During these episodes, symptom severity reached a median VAS of 43 (P25-P75 29-60), with 2% (2/99) of patients reporting emergency department visits and 1% (1/99) an unscheduled medical appointment.

A total of 63% (62/99) of patients answered symptom questionnaires: 62% (61/99) answered at least one CARAT (median 1 [P25-P75 1-2]), 58% (57/99) answered weekly symptom questionnaires (median 4 [P25-P75 2-4]), and 42% (42/99) answered daily symptom questionnaires (median 9.5 [P25-P75 4-13.3]). A total of 76% (75/99) received at least one badge (2 [P25-P75 2-3]), 63% (62/99) achieved 1000 points, and 21% (21/99) achieved 10,000 points. A total of 6% (6/99) of users became Inspirers, a role achieved between 5 and 28 days of app use.

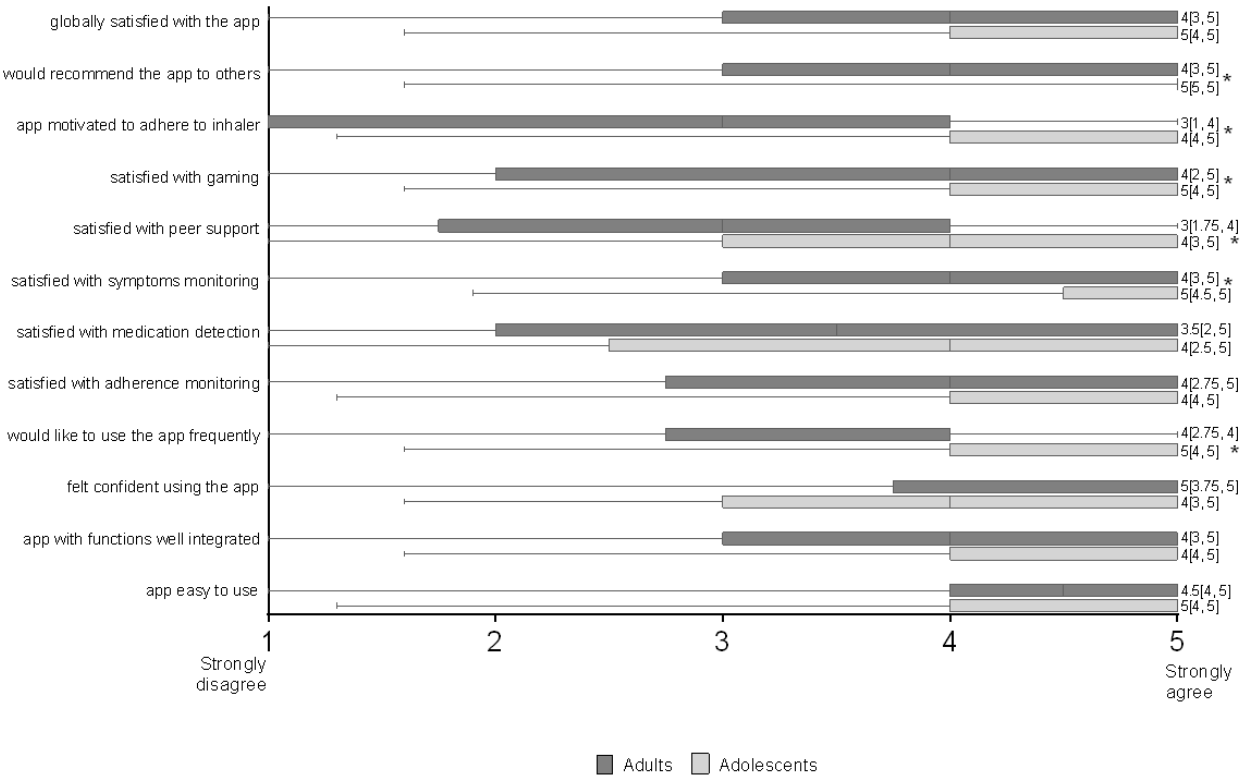
Figure 3. Daily participant engagement with the InspirerMundi app (n=97, 2 users with missing information). Each dot represents a day with interaction, and each color represents a participant.

App Satisfaction

A total of 74% (79/107) of patients completed the 1-month interview (28/107; 26% dropout rate). After 1 month, CARAT scores improved (21 [P25-P75 16-23] vs 25 [P25-P75 22-27]; *P*<.001), with the proportion of patients with controlled asthma increasing from 20% (16/79) to 53% (42/79; *P*<.001). Patients were globally satisfied with the app, with 75% (59/79) scoring ≥4 on a 1 to 5 scale (33% [26/79] scored 4 and 42% [33/79] scored 5), with the adherence monitoring, symptom monitoring, and gamification being the most highly scored components (Figure 4). Only 14% (11/79) were not satisfied with the app (score ≤2). Patients found the app easy to use, considered the functions well integrated, felt confident using the app, and would like to use it frequently. A total of 77% (61/79) of patients would recommend the app to other patients, and 53% (42/79) believed

the app had motivated them to improve adherence to inhaled medication. In general, adolescents were more satisfied with the app than adults (Figure 4).

Figure 4. App feedback from adolescents and adults (n=79). Data are presented as box plots (lines inside the boxes represent the medians; bounds of boxes, first and third quartiles; bars, 95% confidence interval). *Significant differences between adolescents and adults.



App Features Feedback

All patients provided feedback on the app, 95% (75/79) pointed out the aspects that they valued most, 73% (58/79) mentioned aspects least preferred, and 62% (49/79) made improvement suggestions. Patient feedback was reflected in 4 major themes: medication-related features (67/79, 85%), gamification and social network (33/79, 42%), symptom monitoring and physician communication (21/79, 27%), and other aspects (16/79, 20%). These themes are described in detail below.

Medication-Related Features

Patients pointed out the notifications for the medication (47/79, 59%) and adherence statistics (18/79, 23%) as strengths of the app. Yet 2 patients reported problems in receiving notifications, and 2 suggested improvements such as making notifications clearer, sending a second notification if the medication was not taken within 1 hour, and allowing the confirmation of medication intake without entering the app. The available time to confirm the medication intake (2 hours after the notification) was considered inflexible (13/79, 16%) as it did not allow for confirmation if the medication was taken earlier (before the notification) or later. This was seen as a disadvantage as it compromised the game and adherence statistics. Patients asked for the possibility to also confirm medications when outside of scheduled hours (15/79, 19%).

> *[Notifications] It helped me to remember to take the medication.*
>
> *With the adherence statistics, I have a better perception of my real adherence.*
>
> *I lost points because I took the medication out of the scheduled hours, and I was not able to confirm it.*

The need to use the medication detection tool to confirm each medication intake was seen as a weak point (22/79, 28%), mainly due to the predefined time to capture the image, perceived as long. This was especially burdensome for patients with more complex medication regimens (eg, more than one control medication per day). Patients suggested that the medication detection tool could be faster in detecting the dose counter (5/79, 6%), provide feedback of the identified number (2/79, 3%), and be activated only at certain time points (eg, every 5 days or with personalized frequency; 4/79, 5%). Two patients would like the app to include educational content related to asthma and inhalation technique.

> *The time I need to wait for the app to take the picture of the inhaler...*
>
> *There could be an option that allows you not to have to take a photo of the inhaler daily, but to take it, for example, on the first day and after 5 or 7 days.*
>
> *The app could include a demonstration on how to use the inhaler using images.*

Registration of controller or relief medication was reported as difficult by 4 patients and they asked for a simpler process, suggesting, for example, selecting the medication from a provided list, not needing to insert a new medication when a

new inhaler is bought, and personalizing scheduled times depending on the day of the week (eg, during weekends).

> *I found it difficult to register an SOS intake; I need more information to be able to use this feature.*
>
> *When the inhaler is over, it would be nice not to have to register a new inhaler with all the data.*

Gamification and Social Network

Opinions differed with regard to the gamification and social network components: 18% (14/79) of patients (both adolescents and adults) were happy and stated that these components were an additional motivation to improve adherence; 13% (10/79, 9 adults) felt these components were unnecessary or not adding benefit. Two patients reported the gamification approach as confusing, and others did not like the Inspirers avatar.

> *The game increases my motivation to take the medication.*
>
> *I don't have any interest in using the game or to share experiences in the social network.*

Symptom Monitoring and Physician Communication

Patients pointed out the advantages of symptom monitoring (10/79, 13%), with three acknowledging that it increased their perception of disease control (n=3). Three patients also highlighted the utility of the app to share clinical information (use of relief medication, symptoms) with the physician (n=3). One patient suggested a more direct connection with the physician through a chat, for example, to receive recommendations, clarify doubts, and even schedule appointments, and another suggested the addition of features to compare symptom evolution.

> *I liked the symptom questionnaire and interpretation. The questionnaires were simple to fill in.*
>
> *The possibility to have a reliable registration of my disease trajectory and to be able to show to a health professional. I hope to have benefits with this registration!*
>
> *It would be good to chat with my physician, clarify doubts, and schedule appointments.*
>
> *I would like to compare the results of the questionnaires to be able to notice if there was any improvement in symptoms or not.*

Others

Patients reported experiencing technical difficulties/bugs (during log-in, registration of symptom questionnaires, medication detection tool; sometimes the app crashed or closed unexpectedly [8/79, 10%]) and two patients reported that the app used significant storage space (n=2). Other suggestions were a timeline showing the present events by default, simplified log-in (eg, touch ID), and improved access to Definitions and Profile menus.

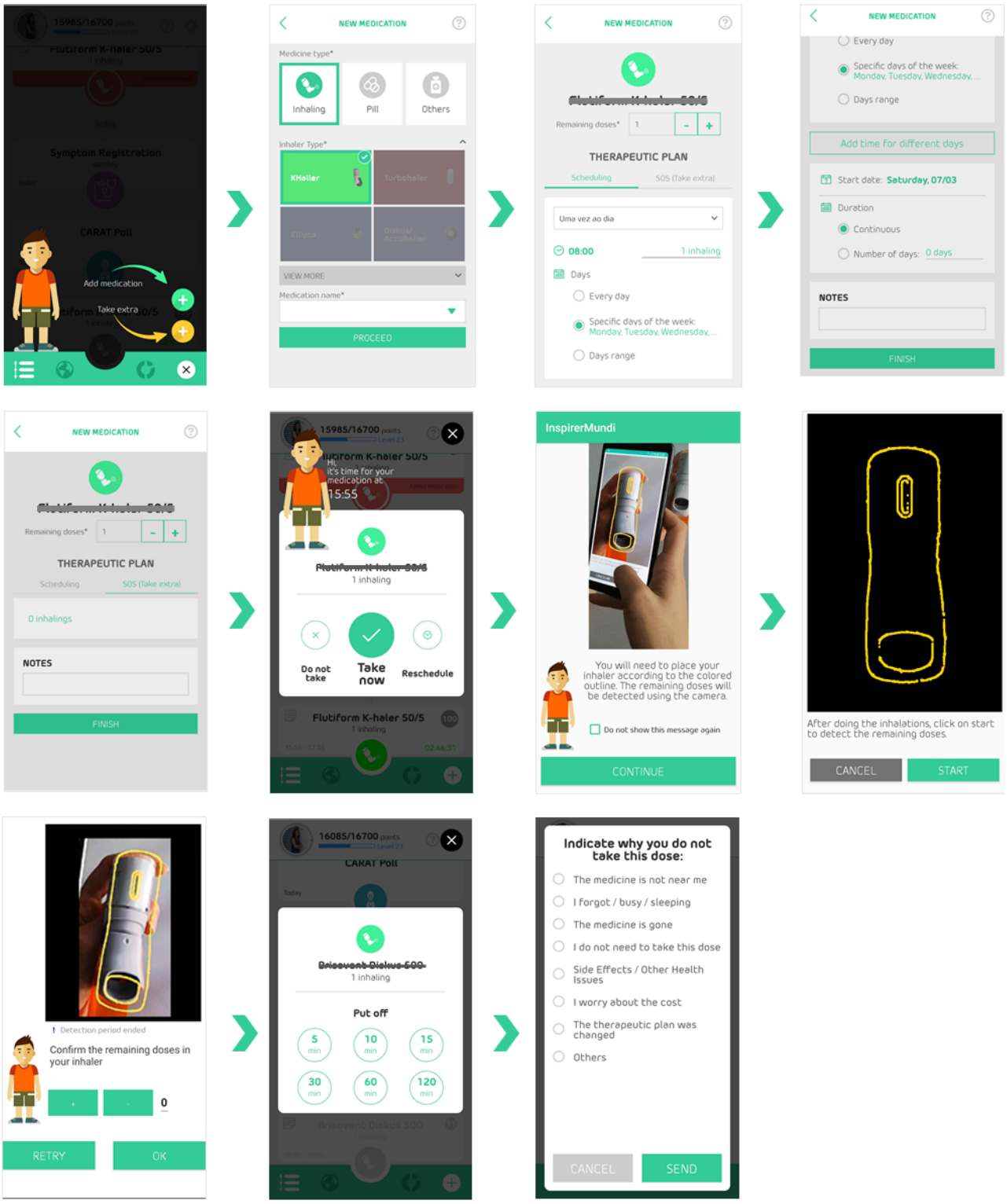
> *The app could be smaller.*
>
> *Log-in could be simplified, and could be with smartphone biometry.*

App Changes Based on Patient Use and Feedback

Based on patient feedback, two major changes were implemented in a new Android app (version 1.2): redesign of the therapeutic plan registration module and medication detection tool. Now, to register an inhaler, the patient is provided with images of the most common inhalers available on the Portuguese and Spanish markets and once the patient selects the inhaler type, a drop-up list presents all names/dosages of medications available on the market. The patient can also select different schedules on different days, allowing the personalization requested. To allow the registration of the complete asthma therapeutic plan, the patient can indicate if the controller inhaler can also be used as relief and how many inhalations are recommended in that case. Also, as registering medications is the only way the app produces notifications, this is now the first task asked once the patient enters the app.

When a medication is due, notifications provide information on the medication and posology and the patient can select 1 of 3 options: confirm medication was taken, delay medication intake (from 5 to 120 minutes), or decide not to take it. Patients now have until the end of the day to confirm medication was taken. The medication detection tool is currently able to detect 12 inhaler devices containing numeric dose counters available on the Portuguese and the Spanish markets. The medication detection tool is activated daily only for inhalers (not for other drug formulations) to reduce the users' burden. When the patient confirms medication was taken, an initial screen with instructions on how to match the inhaler with the template is provided, but in the current version instead of an example figure, we use a short video demonstrating the task (patients can decide to inactivate these instructions whenever they want to accelerate the process). A feature to provide feedback on the identified number in the inhaler was already developed [16] allowing the user to correct the number, and it will be introduced in a future app version. Reasons for deciding not to take a medication can also be registered (eg, medication is empty, this dose is not needed, secondary effects). Screenshots with these app changes are presented in Figure 5, and videos are available [33,34].

Figure 5. InspirerMundi app screenshots (version 1.2).



Discussion

Principal Findings

The InspirerMundi app was found to be feasible to monitor medication adherence and positively evaluated by patients with persistent asthma. The results of this study provide an understanding of the feasibility and acceptability of this mHealth technology in this population. The study contributed to identify improvements for a better user experience before effectiveness studies are undertaken.

Patient use rate of the app is similar to a previous mHealth asthma observational study (median of 12 days in the first 3 months) [35,36], and it is within the range described in digital health studies (median 6 days, range 2 to 26, in the first 3 months) [36]. Nevertheless, we consider the patient use rate lower than expected as the app was directly recommended by a physician and app use was reinforced by 3 text messages and

a telephone interview. This low engagement may be linked to the real-world nature of these studies and the low previous use of asthma apps (<5%) [13]. Importantly, it may also be related to the low level of maturity of some of the app technologies. Less than one-third of patients interacted with the app only 1 day, which is lower than previously described behaviors [36]. These patients were found to be the ones with worst asthma outcomes, such as unscheduled medical visits and exacerbations, and worst perception of health status. Chan et al [35] also observed that the group of patients with fewer interactions with an asthma app were the ones with higher hospitalization rates and emergency department visits. Male patients were also associated with lower interaction, in accordance with previous observations of male patients to be less likely to adopt health apps [37]. These observations need to be considered in the design of mHealth-based interventions to improve adherence to medications.

A small proportion of patients entered the study but did not install the app (8%), which is in line with previous studies with asthma apps [38]. It was not surprising to observe that these patients were older and most commonly were in a first medical appointment. In fact, older patients are the least represented in digital health studies [36,39] and are also less likely to adopt health apps [37]. First appointments with a health care provider are linked to a greater discordance between patients and physicians [40], which may contribute to reducing the known role of physician recommendation as a facilitator of mHealth adoption [36,41]. The total dropout rate (26%) is in accordance with rates in feasibility studies with digital interventions [42] but higher than previously described in clinical trials [38]. This dropout rate was expected given the real-world nature of the feasibility study involving patients from a large number of secondary care centers (26 in total), very diverse in terms of settings—secondary and tertiary care, public and privately owned, dimension, geographic regions, and medical specialties (allergy, pulmonology, and pediatrics). This is contrasted with clinical trials, which are most often conducted in well-selected academic centers with more homogeneous samples and a high number of face-to-face visits.

More than three-quarters of patients used the app for medication management and adherence monitoring, which were the primary aims of app use, and a median score of 4 out of 5 was obtained for this app feature. Self-reported inhaler adherence at the initial visit, 1-week and 1-month interview overestimated adherence measured with the app. This was somewhat expected as self-reports had a near to perfect treatment adherence behavior, which is unlikely to be the case [40]. Overestimation of adherence measured by the app is also possible, as no objective confirmation of the inhaler dose counter was performed in this version. New versions of the app should assess inhaler adherence more objectively by validating inhaler use through dose tracking. The medication detection tool is currently working with 13 inhaler devices containing numeric dose counters available on the Portuguese and Spanish markets [16]. To improve the generalizability of the app and its use in multicenter international studies, inhalers available in other countries will be added in future versions of the app.

Patient evaluations were positive: three-quarters were globally satisfied (75%), reporting scores of 4 (33%) and 5 (42%). The InspirerMundi app acceptability was similar to that described in other feasibility studies of mHealth apps [43-45]. Notifications for medication intake were considered helpful [20,46]. Patients made important contributions for improvement such as the need to make notifications clearer and registration of medications easier. These requested changes are in line with user reviews for other adherence apps [46] and were implemented in the InspirerMundi Android version 1.2, currently being tested by patients recruited at primary care centers. Patients also suggested improvements to the medication detection tool, which led to the tool's redesign, including feedback on the identified number.

Symptom monitoring, used by almost two-thirds of patients, was one of the most highly scored components, with patients stating that it increased their perception of asthma control [20]. Patients valued the feature of sharing information with physician [20,46] and requested other patient-physician partnership tools. In the tested version, patients could send data on the therapeutic plan registered in the app and respective adherence to their physician by email. New simple and easily read reports (including graphs/charts) have been designed, and a tool for remote personalized feedback is being planned. We believe these upgrades will facilitate the integration of app data into clinical workflow and strengthen the patient-physician partnership.

Based on patient feedback, improvements on gamification, social support and monitoring features, and integration of educational content are needed to ensure ongoing digital engagement. The team is considering the integration of new gamification dynamics and mechanics (eg, avatar personalization, progression bars, boosters, exchange points, other tangible rewards), social support features (eg, discussion forums, mentoring, group actions), and educational features (eg, training videos on inhalation technique) [20]. There were opposing opinions regarding the gamification and social network components, also observed in a previous study with adolescents with asthma [20]. To accommodate these differences, and as personalized strategies to treatment adherence have better results [47,48], gamification and social network features of the app will become optional and patients not willing to use these features will be able to turn them off. This modification is in line with previous work advocating that flexible apps may better meet the needs of a broader range of patients [49,50].

Limitations

This study has limitations. As a feasibility study, a control group was not included, and the sample size was not powered. This was a deliberate choice to test the feasibility and acceptability of the app for the first time. As the study was conducted at secondary care centers from Portugal and Spain, results may not be generalized to other settings or countries. An updated version of the app is being tested in Portuguese primary care centers. Participation in the study was restricted to patients who have access to a smartphone. Although smartphones are increasingly prevalent in patients with asthma [13], this requirement excluded otherwise eligible patients and limits the

generalizability to smartphone owners. Using an app to improve treatment adherence will not be a solution for all patients, as happens with most, if not all, approaches. With this study, previous ones, and others planned it should be possible to identify those patients who may benefit from InspirerMundi app strategies. The use of two methods to calculate medication adherence may also be seen as a limitation of the study, yet it was a mitigation plan to overcome the absence of a standardized method to measure adherence [9].

Conclusion

In conclusion, the InspirerMundi app was feasible and acceptable to monitor medication adherence in adolescents and adults with asthma. Based on patient feedback, redesign of the therapeutic plan registration and the medication detection tool have already been implemented, and other changes to the app are being considered. This study is part of a continuous development cycle that will continue with its multiple improvement and evaluation phases grounded on the interaction with end users, aiming to produce a patient-centered and engaging mHealth asthma app.

Acknowledgments

We thank the participants and centers involved in this study. We would also like to acknowledge all members of the Inspirers group [51] and Mundipharma-Portugal for supporting the dissemination of the InspirerMundi app. This work was funded by the European Regional Development Fund through the operations: POCI-01-0145-36 FEDER-029130 (mINSPIRE–mHealth to measure and improve adherence to medication in chronic obstructive respiratory diseases: generalization and evaluation of gamification, peer support, and advanced image processing technologies) and cofunded by COMPETE2020 (Programa Operacional Competitividade e Internacionalização) and Portugal 2020 and with Portuguese funds through Fundação para a Ciência e a Tecnologia.

Conflicts of Interest

JAF is the co-founder of SME, which owns the InspirerMundi app.

References

1. Global strategy for prevention, diagnosis and management of COPD. Global Initiative for Chronic Obstructive Lung Disease. 2020. URL: https://goldcopd.org/wp-content/uploads/2019/12/GOLD-2020-FINAL-ver1.2-03Dec19_WMV.pdf [accessed 2021-05-15]
2. Bender BG. Nonadherence to asthma treatment: getting unstuck. J Allergy Clin Immunol Pract 2016;4(5):849-851. [doi: 10.1016/j.jaip.2016.07.007] [Medline: 27587318]
3. Hassan M, Davies SE, Trethewey SP, Mansur AH. Prevalence and predictors of adherence to controller therapy in adult patients with severe/difficult-to-treat asthma: a systematic review and meta-analysis. J Asthma 2020 Dec;57(12):1379-1388. [doi: 10.1080/02770903.2019.1645169] [Medline: 31311359]
4. Mäkelä MJ, Backer V, Hedegaard M, Larsson K. Adherence to inhaled therapies, health outcomes and costs in patients with asthma and COPD. Respir Med 2013 Oct;107(10):1481-1490 [FREE Full text] [doi: 10.1016/j.rmed.2013.04.005] [Medline: 23643487]
5. van Boven JFM, Lavorini F, Dekhuijzen PNR, Blasi F, Price DB, Viegi G. Urging Europe to put non-adherence to inhaled respiratory medication higher on the policy agenda: a report from the First European Congress on Adherence to Therapy. Eur Respir J 2017 May;49(5):1 [FREE Full text] [doi: 10.1183/13993003.00076-2017] [Medline: 28526801]
6. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency: adherence to medicines for diabetes, hypertension, and hyperlipidaemia. OECD Health Working Papers No 105. Paris: OECD Publishing; 2018. URL: https://www.oecd-ilibrary.org/deliver/8178962c-en.pdf?itemId=%2Fcontent%2Fpaper%2F8178962c-en&mimeType=pdf [accessed 2021-05-15]
7. Kleinsinger F. The unmet challenge of medication nonadherence. Perm J 2018;22:18-033 [FREE Full text] [doi: 10.7812/TPP/18-033] [Medline: 30005722]
8. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
9. Pednekar PP, Ágh T, Malmenäs M, Raval AD, Bennett BM, Borah BJ, et al. Methods for measuring multiple medication adherence: a systematic review-report of the ISPOR Medication Adherence and Persistence Special Interest Group. Value Health 2019 Feb;22(2):139-156 [FREE Full text] [doi: 10.1016/j.jval.2018.08.006] [Medline: 30711058]
10. Howard S, Lang A, Sharples S, Shaw D. What are the pros and cons of electronically monitoring inhaler use in asthma? A multistakeholder perspective. BMJ Open Respir Res 2016;3(1):e000159 [FREE Full text] [doi: 10.1136/bmjresp-2016-000159] [Medline: 27933181]
11. Pereira AM, Jácome C, Almeida R, Fonseca JA. How the smartphone is changing allergy diagnostics. Curr Allergy Asthma Rep 2018 Oct 25;18(12):69. [doi: 10.1007/s11882-018-0824-4] [Medline: 30361774]

12. Haase J, Farris KB, Dorsch MP. Mobile applications to improve medication adherence. Telemed J E Health 2016 Jun 01;23(2):75-79. [doi: 10.1089/tmj.2015.0227] [Medline: 27248315]
13. Jácome C, Almeida R, Pereira A, Araújo L, Correia M, Pereira M, et al. Asthma app use and interest among patients with asthma: a multicenter study. J Investig Allergol Clin Immunol 2020 Apr 23;30(2):137-140 [FREE Full text] [doi: 10.18176/jiaci.0456] [Medline: 32327403]
14. Jácome C, Marques F, Paixão C, Rebelo P, Oliveira A, Cruz J, et al. Embracing digital technology in chronic respiratory care: surveying patients access and confidence. Pulmonology 2020;26(1):56-59 [FREE Full text] [doi: 10.1016/j.pulmoe.2019.05.001] [Medline: 31160235]
15. Jácome C, Almeida R, Teixeira J, Vieira-Marques P, Vilaça R, Fernandes J, et al. Inspirers: an app to measure and improve adherence to inhaled treatment. In: IADIS International Conference e-Health 2017 (part of MCCSIS 2017).: IADIS Press; 2017 Presented at: IADIS International Conference e-Health; 2017; Lisbon p. 135-139 URL: http://www.iadisportal.org/digital-library/inspirers-an-app-to-measure-and-improve-adherence-to-inhaled-treatment
16. Vieira-Marques P, Almeida R, Teixeira JF, Valente J, Jácome C, Cachim A, et al. InspirerMundi: remote monitoring of inhaled medication adherence through objective verification based on combined image processing techniques. Methods Inf Med 2021 Apr 27:1. [doi: 10.1055/s-0041-1726277] [Medline: 33906260]
17. Ferreira A, Almeida R, Jácome C, Fernandes J, Fonseca J, Vieira-Marques P. How inspiring is your app: a usability take on an app for asthma medication adherence. In: IADIS International Conference e-Health 2019 (part of MCCSIS 2019).: IADIS Press; 2019 Presented at: Proceedings of the 11th International Conference e-Health; 2019; Lisbon p. 225-229. [doi: 10.33965/eh2019_201910c030]
18. Jácome C, Guedes R, Almeida R, Teixeira J, Pinho B, Vieira-Marques P, et al. [mINSPIRERS: Feasibility of a mobile application to measure and improve adherence to inhaled controller medications among adolescents and adults with persistent asthma. Protocol for a multicentre observational study]. Rev Port de Imunoalergologia 2018;26(1):47-61 [FREE Full text] [doi: 10.14488/enegep2017_tn_sto_238_376_34297]
19. Billingham SAM, Whitehead AL, Julious SA. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. BMC Med Res Methodol 2013;13:104 [FREE Full text] [doi: 10.1186/1471-2288-13-104] [Medline: 23961782]
20. Schneider T, Baum L, Amy A, Marisa C. I have most of my asthma under control and I know how my asthma acts: users' perceptions of asthma self-management mobile app tailored for adolescents. Health Informatics J 2019 Feb 08;26(1):342-353. [doi: 10.1177/1460458218824734] [Medline: 30732520]
21. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg 2014 Dec;12(12):1500-1524 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.014] [Medline: 25046751]
22. Fogg B. A behavior model for persuasive design. 2009 Presented at: Proceedings of the 4th International Conference on Persuasive Technology; 2009; Claremont p. 40. [doi: 10.1145/1541948.1541999]
23. Fonseca JA, Nogueira-Silva L, Morais-Almeida M, Azevedo L, Sa-Sousa A, Branco-Ferreira M, et al. Validation of a questionnaire (CARAT10) to assess rhinitis and asthma in patients with asthma. Allergy 2010 Aug;65(8):1042-1048. [doi: 10.1111/j.1398-9995.2009.02310.x] [Medline: 20121755]
24. Fonseca JA, Nogueira-Silva L, Morais-Almeida M, Sa-Sousa A, Azevedo LF, Ferreira J, et al. Control of Allergic Rhinitis and Asthma Test (CARAT) can be used to assess individual patients over time. Clin Transl Allergy 2012 Aug 30;2(1):16 [FREE Full text] [doi: 10.1186/2045-7022-2-16] [Medline: 22935298]
25. Azevedo P, Correia DSJ, Bousquet J, Bugalho-Almeida A, Del Giacco SR, Demoly P, WHO Collaborative Center for AsthmaRhinitis, Montpellier. Control of Allergic Rhinitis and Asthma Test (CARAT): dissemination and applications in primary care. Prim Care Respir J 2013 Mar;22(1):112-116 [FREE Full text] [doi: 10.4104/pcrj.2013.00012] [Medline: 23412110]
26. InspirerMundi version 1.1 inhaler registration. URL: https://www.youtube.com/watch?v=wZ5yRhpzSbo [accessed 2021-05-16]
27. InspirerMundi version 1.1 intake registration. URL: https://www.youtube.com/watch?v=NTuOKJ1ZEtc [accessed 2021-05-16]
28. InspirerMundi version 1.1 CARAT questionnaire. URL: https://www.youtube.com/watch?v=9EXsZouE9Ro [accessed 2021-05-16]
29. InspirerMundi version 1.1 Inspirer request. URL: https://www.youtube.com/watch?v=8Q72VWVH5pU [accessed 2021-05-16]
30. Loymans RJB, Ter Riet G, Sterk PJ. Definitions of asthma exacerbations. Curr Opin Allergy Clin Immunol 2011 Jun;11(3):181-186. [doi: 10.1097/ACI.0b013e3283466f45] [Medline: 21467926]
31. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee TA. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. Respir Med 2008 Apr;102(4):519-536 [FREE Full text] [doi: 10.1016/j.rmed.2007.11.016] [Medline: 18180151]
32. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese Validation of the System Usability Scale (SUS). Procedia Comput Sci 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]
33. InspirerMundi version 1.2 inhaler registration. URL: https://www.youtube.com/watch?v=OlI_ln3RPbs [accessed 2021-05-16]
34. InspirerMundi version 1.2 intake registration. URL: https://youtu.be/E2qP2enArqE [accessed 2021-05-16]

35. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Mar 13;35(4):354-362. [doi: 10.1038/nbt.3826] [Medline: 28288104]
36. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
37. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
38. Kosse RC, Bouvy ML, Belitser SV, de Vries TW, van der Wal PS, Koster ES. Effective engagement of adolescent asthma patients with mobile health-supporting medication adherence. JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12411 [FREE Full text] [doi: 10.2196/12411] [Medline: 30916664]
39. Neves A, Jácome C, Taveira-Gomes T, Pereira AM, Almeida R, Amaral R, et al. Determinants of use of health and fitness mobile apps by patients with asthma: a secondary analysis of observational studies. J Med Internet Res 2021:1 (forthcoming) [FREE Full text] [doi: 10.2196/25472]
40. Jácome C, Pereira AM, Almeida R, Ferreira-Magalhaes M, Couto M, Araujo L, Inspirers group. Patient-physician discordance in assessment of adherence to inhaled controller medication: a cross-sectional analysis of two cohorts. BMJ Open 2019 Nov 07;9(11):e031732 [FREE Full text] [doi: 10.1136/bmjopen-2019-031732] [Medline: 31699737]
41. VonHoltz LAH, Hypolite KA, Carr BG, Shofer FS, Winston FK, Hanson CW, et al. Use of mobile apps: a patient-centered approach. Acad Emerg Med 2015 Jun;22(6):765-768 [FREE Full text] [doi: 10.1111/acem.12675] [Medline: 25998446]
42. Merchant R, Inamdar R, Henderson K, Barrett M, Su JG, Riley J, et al. Digital health intervention for asthma: patient-reported value and usability. JMIR Mhealth Uhealth 2018 Jun 04;6(6):e133 [FREE Full text] [doi: 10.2196/mhealth.7362] [Medline: 29866644]
43. Hoogland J, Wijnen A, Munsterman T, Gerritsma CL, Dijkstra B, Zijlstra WP, et al. Feasibility and patient experience of a home-based rehabilitation program driven by a tablet app and mobility monitoring for patients after a total hip arthroplasty. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e10342 [FREE Full text] [doi: 10.2196/10342] [Medline: 30702438]
44. Davidson S, Fletcher S, Wadley G, Reavley N, Gunn J, Wade D. A mobile phone app to improve the mental health of taxi drivers: single-arm feasibility trial. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e13133 [FREE Full text] [doi: 10.2196/13133] [Medline: 31939743]
45. Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]
46. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
47. Etminani K, Tao Engström A, Göransson C, Sant'Anna A, Nowaczyk S. How behavior change strategies are used to design digital interventions to improve medication adherence and blood pressure among patients with hypertension: systematic review. J Med Internet Res 2020 Apr 09;22(4):e17201 [FREE Full text] [doi: 10.2196/17201] [Medline: 32271148]
48. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
49. Hui CY, Walton R, McKinstry B, Pinnock H. Time to change the paradigm? A mixed method study of the preferred and potential features of an asthma self-management app. Health Informatics J 2019 Jun 13:1460458219853381. [doi: 10.1177/1460458219853381] [Medline: 31195877]
50. Hui C, Walton R, McKinstry B, Pinnock H. What features do patients "want" in a mobile asthma app? A qualitative study. Eur Respir J 2017;50(suppl 61):PA3874. [doi: 10.1183/1393003.congress-2017.PA3874]
51. Inspirers group. URL: https://paceit.med.up.pt/sample-page/inspirers-group/ [accessed 2021-05-16]

Abbreviations

CARAT: Control of Allergic Rhinitis and Asthma Test
EQ-5D: EuroQol 5-Dimension questionnaire
GINA: Global Initiative for Asthma
mHealth: mobile health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VAS: visual analog scale

Edited by L Buis; submitted 14.12.20; peer-reviewed by A(Kassavou, HL Tam; comments to author 11.01.21; revised version received 26.01.21; accepted 09.04.21; published 25.05.21.

Please cite as:
Jácome C, Almeida R, Pereira AM, Amaral R, Mendes S, Alves-Correia M, Vidal C, López Freire S, Méndez Brea P, Araújo L, Couto M, Antolín-Amérigo D, de la Hoz Caballer B, Barra Castro A, Gonzalez-De-Olano D, Todo Bom A, Azevedo J, Leiria Pinto P, Pinto N, Castro Neves A, Palhinha A, Todo Bom F, Costa A, Chaves Loureiro C, Maia Santos L, Arrobas A, Valério M, Cardoso J, Emiliano M, Gerardo R, Cidrais Rodrigues JC, Oliveira G, Carvalho J, Mendes A, Lozoya C, Santos N, Menezes F, Gomes R, Câmara R, Rodrigues Alves R, Moreira AS, Bordalo D, Alves C, Ferreira JA, Lopes C, Silva D, Vasconcelos MJ, Teixeira MF, Ferreira-Magalhães M, Taborda-Barata L, Cálix MJ, Alves A, Almeida Fonseca J
Feasibility and Acceptability of an Asthma App to Monitor Medication Adherence: Mixed Methods Study
JMIR Mhealth Uhealth 2021;9(5):e26442
URL: https://mhealth.jmir.org/2021/5/e26442
doi:10.2196/26442
PMID:34032576

Original Paper

Smartphone Usage Among Doctors in the Clinical Setting in Two Culturally Distinct Countries: Cross-sectional Comparative Study

Anjali Ajay Nair[1*], MB BCH BAO; Samreen Afroz[1*], MB BCH BAO; Bushra Urooj Ahmed[1*], MB BCH BAO; Uzma Urooj Ahmed[1*], MB BCH BAO; Chi Chung Foo[2*], MBBS, MS; Hind Zaidan[3,4*], MD; Martin Corbally[3*], MB BCH BAO, MCh, MD

[1]School of Medicine, RCSI Bahrain, Busaiteen, Bahrain

[2]Department of Surgery, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Hong Kong

[3]Department of General Surgery, King Hamad University Hospital, Busaiteen, Bahrain

[4]Our Lady's Children's Hospital, Crumlin, Dublin, Ireland

[*]all authors contributed equally

Corresponding Author:
Anjali Ajay Nair, MB BCH BAO
School of Medicine
RCSI Bahrain
Building No. 2441
Road 2835
Busaiteen
Bahrain
Phone: 852 6759 5753
Email: 14034735@rcsi.com

Abstract

Background: Smartphones and mobile applications have seen a surge in popularity in recent years, a pattern that has also been reflected in the health care system. Despite increased reliance among clinicians however, limited research has been conducted on the uptake and impact of smartphone usage in medical practice, especially outside the Western world.

Objective: This study aimed to identify the usage of smartphones and medical apps by doctors in the clinical setting in 2 culturally distinct countries: King Hamad University Hospital (KHUH), Bahrain and Queen Mary Hospital (QMH), Hong Kong.

Methods: A cross-sectional, comparative study was conducted where doctors in both hospitals were asked to take part in a 15-item online survey. The questions were categorized into the following groups: demographics of the study population, ownership and main use of smartphones, number and names of medical apps currently owned, rating usage of smartphones for medical purposes, time spent on a smartphone related to clinical use, clinical reliance on smartphones, and views on further integration of smartphones. The results were then tabulated and analyzed using SPSS Statistics 25 for Mac (IBM Corp Inc, Armonk, NY).

Results: A total of 200 doctors were surveyed, with a total of 99.0% (99/100) of the doctors owning a smartphone in both KHUH and QMH; 58% (57/99) and 55% (54/99) of the doctors from KHUH and QMH, respectively, identified communication as their main use of smartphones in the clinical setting (P=.004). Doctors from KHUH were likely to spend more time on medical apps than doctors from QMH (P=.002). According to the overall results of both hospitals, 48% (32/67) of the junior doctors claimed high reliance on smartphones, whereas only 32.3% (41/127) of the senior doctors said the same (P=.03). Of doctors in KHUH and QMH, 78.0% (78/100) and 69.0% (69/100), respectively, either strongly agreed or agreed that smartphones need to be integrated into the clinical setting. In terms of preferences for future apps, 48% (48/100) and 56% (56/100) of the doctors in KHUH and QMH, respectively, agreed that more medical applications need to be created in order to support smartphone use in the clinical setting.

Conclusions: These results suggest a substantial acceptance of smartphones by doctors in the clinical setting. It also elicits the need to establish policies to officially integrate smartphone technology into health care in accordance with ethical guidelines. More emphasis should be placed on creating medical applications that aid health care professionals in attaining their information from accurate sources and also regulate a system to monitor the usage of mobile devices within hospitals to prevent a breach of patient privacy and confidentiality.

(JMIR Mhealth Uhealth 2021;9(5):e22599) doi:10.2196/22599

KEYWORDS

smartphone use; mobile phone; mobile technology; smartphone technology; medical apps; mobile applications; smartphone applications; mHealth; mobile health; digital health; medical informatics; internet; doctors; patient care; point of care; Bahrain; Hong Kong

Introduction

Mobile Health and Smartphones in Health Care

Innovation is one of the hallmarks of medical progress, but traditionally physicians are, as a group, reluctant to embrace all aspects of technical change. As a knowledge-based specialty, the practice of medicine has constantly evolved to ensure wider, faster, and more accurate access to information. Information sharing has changed the way we complete research and share that information for the benefit of our patients. There remains, however, a reluctance for doctors to immerse themselves fully in or on this information highway, and often, this reluctance is based on concerns of confidentiality and a passive acceptance of established practice. Smartphones have had a major impact on rapid access to information and have made major inroads into the mechanisms of information transfer for the profession in general [1]. By and large, smartphones are increasingly used as a means to improve communication between practitioners, keep up to date with the latest medical news and practices, check up on patients post-discharge, look up reference values, and differentials and drug prescriptions as well as provide dynamic education for doctors [1-5]. Studies roughly estimate that 98.4% of doctors own a smartphone and 92% agreed that smartphones positively impacted their practice [2,5-7].

Mobile health (mHealth) is a term used to describe the advances in mobile applications and associated innovations in technology to aid patient care. Being a major constituent of mHealth, mobile devices have drastically evolved since their first introduction in 1973 [1,8]. Doctors are now able to view medical textbooks and calculate drug formulas through apps, eliminating human errors and thus saving time [8,9]. A study on mHealth published in 2017 in Turkey found that communication and information gathering were the major uses of smartphones among physicians [3,10]. With the introduction of smartphones into hospitals, there has also been, peri passu, a simultaneous increase in the demand for medical applications, which has ushered in a new era of convenience. There are currently over 40,000 medical apps available for download in various app stores, and this figure will only continue to rise with increasing popularity [4,11].

Aim of the Study

Despite this increased dependency, limited research has been conducted on the impact of smartphone usage and how it can change medical practice. While there are some data from the United Kingdom and other Western countries, there remains a lack of research in the Middle East and Asia. In this study, we aimed to identify the usage of smartphones and medical apps by doctors in the clinical setting in 2 culturally distinct countries: King Hamad University Hospital (KHUH), Kingdom of Bahrain and Queen Mary Hospital (QMH), Hong Kong. We attempted to understand how the socioeconomic status and culture of the 2 countries impacts the views of health professionals towards smartphones and their use and acceptance in the clinical setting.

Methods

Study Design

We conducted a cross-sectional study over a period of 2 weeks in January 2018 in KHUH, Kingdom of Bahrain and QMH, Hong Kong, both of which are tertiary referral centers and academic hospitals. We designed an online questionnaire on Google Forms based on results from a preliminary literature search [6,7,10]. The questionnaire (which can be found in Multimedia Appendix 1 and Multimedia Appendix 2) consisted of 15 items that were categorized into the following groups: demographics of the study population, ownership and main use of smartphones, number and names of medical apps currently owned, rating usage of smartphones for medical purposes, time spent on a smartphone related to clinical use, clinical reliance on smartphones, and views on further integration of smartphones.

Study Participants

The questionnaire was open to all doctors irrespective of grade or specialty. To make a fair comparison, we decided to assess the same 15 departments in both hospitals with the responses from the additional departments marked as “others.” The exclusion criteria were medical students, nurses, patients, and hospital administrative staff.

Data Collection and Statistical Analysis

We set up a booth in an open forum to allow doctors to approach the research investigators. The set-up was kept identical in both hospitals. Participants were given a brief description and purpose of the research, and consent was implied by their agreement to take part in the study. No defining personal information was recorded, and researchers were present to answer any questions raised. Participants completed the questionnaires in person through laptops or iPads provided by the researchers. The questionnaire was administered on Google Forms, and the responses were then downloaded into a Microsoft Excel sheet (Microsoft Excel For Mac Version 16.10.0, Build 18021001, Microsoft Corp, Redmond, WA). We kept the data acquired from the study in an encrypted folder on the principal investigator’s laptop (to remain for 5 years, after which it will be permanently deleted). The results were tabulated and analyzed using SPSS Statistics 25 for Mac (IBM Corp, Armonk, NY).

Ethics Approval

The study protocol was reviewed and approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (ethics

approval number: UW 17-451); KHUH, Bahrain (ethics approval number: KHUH/Research/No. 179/2017); and the Royal College of Surgeons in Ireland Research Ethics Committee (ethics approval number: RCSIBAH15012018).

Results

The responses from 200 doctors, 100 from each hospital, were recorded, and statistical analysis was performed using SPSS to enable the comparison between the 2 hospitals. The results were tabulated under the headings outlined in the following sections.

Demographics

A similar demographic pattern was observed, with 24.0% (24/100) of the doctors who participated in the study in each hospital belonging to the age group of 26-30 years, with the same male predominance of 67.0% (67/100) in both the hospitals.

The predominant medical specialty in KHUH was general medicine, with an 18.0% (18/100) response rate, whereas 32.0% (32/100) of the doctors from the department of surgery formed the major specialty who responded in QMH, as shown in Table 1. The demographics of the respondents are shown in Multimedia Appendix 3.

A total of 99.0% (99/100) of the doctors in both KHUH and QMH owned a smartphone. From those who owned a smartphone, 58% (57/99) and 55% (54/99) of the doctors from KHUH and QMH, respectively, identified communication as their main use of smartphones in the clinical setting (*P*=.004; Multimedia Appendix 4; Table 2).

Table 1. Medical specialty of the participants in King Hamad University Hospital (KHUH) and Queen Mary Hospital (QMH).

Medical specialty	KHUH (n=100), n	QMH (n=100), n
Accident and emergency	5	1
Anesthetics and intensive care unit (ICU)	14	17
Ear, nose, and throat (ENT)	1	0
General medicine	18	0
Oncology	0	3
General surgery	13	32
Gynecology	5	5
Medicine	0	17
Neurosurgery	1	4
Ophthalmology	3	0
Oral and maxillofacial	1	0
Orthopedic	9	7
Pediatric	11	4
Pathology	0	4
Radiology	5	4
Others	14	2

Table 2. Smartphone use by doctors in King Hamad University Hospital (KHUH) and Queen Mary Hospital (QMH).

Main use of smartphones	KHUH (n=99), n	QMH (n=99), n
Search engines	8	23
Camera	1	0
Communication	57	54
Viewing patient information	4	0
Radiology films	0	0
Drug formulas	2	0
Personal use	22	22
All of the above	5	0

Medical App Ownership

Of the doctors from KHUH and QMH who owned a smartphone, 87% (86/99) and 86% (83/97), respectively, downloaded medical apps; 63% (58/92) of the doctors in KHUH and 51% (48/95) of the doctors in QMH owned 1-3 medical apps, as Table 3 shows.

Medscape and UpToDate were the most used medical apps in KHUH, with 68% (60/88) of the doctors opting for each, while Medscape was the major app used by 68% (57/84) of the doctors in QMH.

Table 3. Medical app ownership by doctors in King Hamad University Hospital (KHUH) and Queen Mary Hospital (QMH).

Number of medical apps owned	KHUH (n=92), n	QMH (n=95), n
0	5	11
1-3	58	48
4-5	16	22
≥6	13	14

Clinical Reliance on Smartphones

The clinical reliance on the apps was analyzed based on different means of attaining information, which involved the following categories: review medical news, hospital information system (online record of patient labs and reports), drug related, communication with patient, communication with colleagues, teaching purposes, training purposes, research purposes, patient education, patient monitoring, and continuing medical education activities.

When asked about the time spent using medical apps in the clinical setting, 29% (29/99) of the total respondents from KHUH used the apps for more than 2 hours per day, 68% (67/99) used apps for 2 or fewer hours, and 3% (3/99) of the doctors never used medical apps, whereas in QMH, 10% (10/97) used apps for more than 2 hours, 83% (80/97) used apps for 2 or fewer hours, and 7% (7/97) never used medical apps in the clinical setting. This is illustrated in Table 4. Doctors from KHUH were likely to spend more time on medical apps in the clinical setting. This difference was significant (P=.002, between the 3 time categories).

Table 4. Time spent on a smartphone by doctors in King Hamad University Hospital (KHUH) and Queen Mary Hospital (QMH).

Time spent on a smartphone	KHUH (n=99), n	QMH (n=97), n
Never use a smartphone in the clinical setting	3	7
≤2 hours	67	80
>2 hours	29	10

The clinical reliance of doctors on smartphones was recorded on a scale of 1 to 5 (5 being maximum and 1 being minimum). The ratings of 1 and 2 were considered “Low,” 3 was considered as “Neutral,” whereas 4 and 5 were considered “High.”

This was compared with the seniority of the doctors in both the hospitals, where Interns or House Officers and Senior House Officers or Medical Officers were grouped as junior doctors, while Residents, Registrars, or Senior Registrars and Consultants, Associate Professors, or Professors were grouped as senior doctors.

Of the junior doctors in KHUH, 57% (24/42) claimed to be highly reliant on their smartphones in the clinical setting, whereas only 29% (16/55) of the senior doctors said the same. The data at KHUH were significant (P=.005).

However, 32% (8/25) of the junior doctors were highly reliant on their medical apps, whereas 35% (25/72) of the senior doctors rated their clinical reliance on smartphones as High in QMH. These data were insignificant (P=.81).

According to the overall results in both the hospitals, a significant difference was seen, where 48% (32/67) of the junior doctors claimed high reliance on smartphones, whereas only 32.3% (41/127) of the senior doctors said the same (P=.03).

Results for Future Integration of Smartphones

It was observed that 90.0% (90/100) and 84.0% (84/100) of the doctors in KHUH and QMH, respectively, “Agreed” or “Strongly Agreed” that smartphones have a huge potential in the clinical setting, and 78.0% (78/100) of the participants in KHUH and 69.0% (69/100) of the doctors in QMH “Strongly Agreed” or “Agreed” that smartphones should be formally integrated into the clinical setting.

In addition, 48.0% (48/100; KHUH) and 56.0% (56/100; QMH) agreed that more medical apps need to be created in order to support smartphone use in the clinical setting, and 38.0% (38/100; KHUH) and 55.0% (55/100; QMH) agreed that they would use their smartphones more if there were more medical apps available for use.

Future Preferences

When asked about the kind of apps they would like to use in the future, drug-related apps formed the major preference for doctors at KHUH (66/100, 66.0%), whereas hospital information systems was the major preference in QMH (65/100, 65.0%), as demonstrated in Table 5.

Table 5. Types of future apps preferred by doctors in King Hamad University Hospital (KHUH) and Queen Mary Hospital (QMH). Participants were asked to select up to 3 choices.

Types of preferred future apps	KHUH (n=100), n	QMH (n=100), n
Review medical news	40	44
Hospital information systems	54	65
Drug related	66	64
Communication with patients	24	14
Communication with colleagues	28	31
Teaching purposes	41	17
Training purposes	34	25
Research purposes	38	23
Patient education	0	0
Continuing medical education (CME) activities	37	13
None	3	2
Other	0	0

Discussion

The study indicates a promising shift in the use of smartphones in the clinical setting. Our study is unique because, to our knowledge, it is one of the only studies that compares the uptake of smartphones of 2 hospitals that come from 2 different cultural backgrounds. It also allows for direct comparison to data collected from Western countries. For example, a study was conducted in the United Kingdom in 2015 where researchers assessed the usage of smartphones among surgeons. Participants from all levels of training were also asked to complete a 13-item questionnaire that assessed their smartphone use in the workplace. Results showed that 93.5% (319/341) owned a smartphone and 54.2% (173/319) within that group also owned medical apps; 79.3% (253/319) stated they would be willing to use their smartphones for hospital-based work [12]. In comparison, our study found higher values of smartphone ownership (99/100, 99.0% in both hospitals) and medical app usage (86/99, 87%, KHUH; 83/97, 86%, QMH), perhaps since the 2 studies were conducted 3 years apart; the later timeframe might have contributed to the differences in values obtained due to the continuous advancements in the field of smartphone technology [5].

Doctors from both countries expressed an interest in the integration of smartphones in the clinical setting, with the majority preferring to see more drug-related apps and hospital information systems in the future. This preference, if accurately employed, could greatly serve to increase efficiency and safety in both drug prescription and accessing patient information, thus saving valuable time.

Compared to previous literature, this study has emphasized a rise in the number of doctors who are clinically reliant on their smartphone. The study, true to its objectives, presents an understanding of the acceptance of technology in the 2 distinct regions.

Greater Smartphone Reliance Among Junior Doctors Than Senior Doctors

Junior doctors in KHUH were more reliant on smartphones and medical apps than their senior colleagues. However, in QMH, senior doctors were more reliant. This may represent the higher number of senior doctors in QMH (Multimedia Appendix 3).

The results showed that more than half of the junior doctors in KHUH said they were highly reliant on smartphones, whereas the majority of the senior doctors in KHUH asserted low clinical reliance. However, the results in QMH insignificantly varied, with senior doctors being more dependent on medical apps. This could be attributed to the fact that 74% (73/99) of the respondents who participated in QMH were senior doctors, and this could have presented bias in the results.

The results obtained from KHUH are in accordance with a UK-based study done in 2015 within the surgical profession. It was found that junior doctors were more up to date with technology and hence were more clinically reliant on medical apps as compared to their senior colleagues [12]. A similar study conducted by Samsung Medical Center reported that 83.3% of the total access of information via smart devices was also done by junior physicians [13].

With easy and flexible access to medical apps, it has vast potential to enhance current medical practice by attaining fast and comprehensive knowledge, thereby improving the outcome of clinical decisions [4].

Implications of Smartphone Usage in the Clinical Setting

The majority of participants from both hospitals agreed that smartphones have a future in clinical practice and that doctors would use their smartphones more if there were more apps created. This illustrates a genuine interest from clinicians to optimize smartphone technology in the workplace if (1) smartphones are formally integrated into the clinical setting and (2) more medical apps are created that support smartphone use

in medical practice. As a result, the implications can be further categorized based on the different stakeholders involved.

Doctors as Users of Smartphones in the Clinical Setting

There is no doubt that smartphones are widely appreciated and used by physicians. Although still emerging in medicine, there have been several reports where doctors have started to officially integrate smartphones into clinical practice. For example, the study conducted at Samsung Medical Center demonstrated how introducing a hospital-specific smartphone app, Dr. SMART S, helped doctors access information regarding inpatients and outpatients (such as lab results) and consultation notes anywhere within the hospital. The app received praise from doctors due to its ease of use and the ability to increase interdepartmental communication [13]. Medical apps have also been slowly introduced in the clinical setting in recent years. In 2015, another study conducted in the United Kingdom assessed how newly qualified doctors used iDoc, an app with medical textbooks that are vital to the doctors' learning. The results revealed a shift away from hardcopy and electronic textbooks to smartphone textbooks; out of a total of 125 participants, over half said they used the app daily for 12 months. In addition, it was found that the majority of the participants also felt that there was a place for smartphones in the workplace, results that are in line with our research [3].

On a user level, it is also important to consider the factors that can influence the doctor's experience with and thus their clinical decision making on the smartphone. Special attention should be paid to the design of the app (interface design, app performance, and cost), context within which the interaction occurs, user's lifestyle, and appropriateness of the activity [14]. Understanding these factors that influence the user experience is an important area of focus for future research of mHealth interventions as they can potentially increase uptake and acceptability of interventions.

Monitoring and Regulation of Usage by Hospitals or Governments

There is a huge demand for medical-related apps for health care professionals from many different specialties due to the potential of smartphones to improve efficiency in a variety of aspects in the medical field [15,16]. These advances include, but are not limited to, increased accessibility, portability or mobility, and communicability [2]. In turn, this would help improve the management of patient records, access to online resources, patient monitoring, access to educational aids, and communication among medical staff, all of which contribute significantly to better patient care and hence patient outcomes [13,17,18].

Despite its many possible benefits, smartphone technology also brings to light negative implications that need to be explored before its full integration into health care. One of the major concerns with this type of technology is the potential for a breach of patient confidentiality. To combat this growing problem however, hospitals, where smartphone devices are already employed, have introduced security systems that manage and monitor devices being used within hospital perimeters and also allow for disabling of devices that are not in use anymore [19]. Furthermore, several countries have adopted regulations to step up measures to protect patient information. In the United States, for example, there are several features smartphones must have to comply with the regulations stipulated by the Personal Health Information Protection Act. These include, but are not limited to, the following: encryption of patient information being transmitted along secured networks, password protection of devices, and automatic wiping of data that are no longer required [20]. Similarly, the General Data Privacy Regulation is an initiative that has been established in the European Union that offers strict laws on the management of mHealth that institutions have to adhere to [21].

Study Limitations

As with the majority of studies, the design of the study is subject to limitations. First, it was predecided — keeping in mind the number of doctors currently in QMH and KHUH — that the study will be concluded after receiving 100 responses in each hospital, with a total of 200 participants. Despite the small sample size, we believe our results are representative of the survey population due to the high response rate and lack of selection bias inherent in an online survey, which also ensured consistency. A possible bias of our study is social desirability bias. The participants, while completing the survey, could have responded in a way that portrayed them in a positive light or would be viewed favorably to us. We tried to minimize this by keeping the participants anonymous so as to remove undue pressure and allow them to respond freely.

Our survey depended on the doctor's ability to recall past events, which could have introduced recall bias. We took this into account prior to conducting the study by devising a high-quality questionnaire, allowing participants sufficient time for adequate recall, and being present physically at the booth to answer questions they might have [22].

Finally, although participants were selected randomly based on opportunity, it was observed that 73.7% of the doctors recruited in QMH were seniors (Residents, Associate Professors, Professors, or Consultants), making the cohort of senior doctors considerably greater than that of junior doctors. We believe this could be due to the fact they were more willing to participate as they were not as busy as junior doctors. However, as senior doctors form an essential part of the clinical setting in Hong Kong, we chose to include their responses in our study.

Conclusion

Interestingly, despite cultural differences between Bahrain and Hong Kong, there were no significant differences noted between the results obtained from the 2 countries. These results suggest a substantial acceptance of smartphone technology by doctors in clinical settings. Hence, more emphasis should be placed on creating medical apps that support doctors in patient care, especially for drug-related uses and hospital information systems, which were found to be the major preferences for future apps by doctors in this study.

However, with an increase in the usage of smartphones in the medical field, there are growing concerns about the protection of patient information. There is a need to establish policies to officially integrate the technology in accordance with ethical

guidelines and thus encourage future studies to discuss guidelines that are patient-centered and respect patient privacy. This will aid health care professionals in attaining their information from accurate sources and also regulate a system to monitor the usage of mobile devices within hospitals to prevent the breach of patient privacy and confidentiality. Though challenging, implementation of evidence-based guidelines governing smartphone use as well as limiting access to patient information outside of the hospital can help overcome these issues [3,21]. Nevertheless, it is clear through our research and the increased number of collaborative studies on the topic in the last decade [23] that the potential for mobile communication to transform health care and clinical intervention in the community is tremendous [24]. There is, therefore, great scope to harness the potential of mobile use in the clinical setting to improve several aspects of health care [23]. Future studies should aim to expand on the existing research by exploring different contexts, especially ones that compare multiple different contexts such as the current study. This will help to gain a better understanding on whether culture can influence smartphone uptake in the clinical setting.

Acknowledgments

The authors would like to thank Ms Abeer, Research Officer in the School of Postgraduate Studies and Research at RCSI Bahrain, for her support and guidance in the initial phase of this research study. We would also like to thank Dr Salim Fredricks, Senior Lecturer in Biochemistry at RCSI Bahrain, for sharing his expertise in statistical analysis with us and providing guidance on how to analyze our data.

In addition, we would also like to thank administrative staff at both KHUH, Bahrain and QMH, Hong Kong for granting permission to use their venues and assisting in the set up of the study.

Finally, we would like to thank RCSI Bahrain for funding the publication of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
KHUH questionnaire.
[PDF File (Adobe PDF File), 127 KB - mhealth_v9i5e22599_app1.pdf]

Multimedia Appendix 2
QMH questionnaire.
[PDF File (Adobe PDF File), 171 KB - mhealth_v9i5e22599_app2.pdf]

Multimedia Appendix 3
Table S1: Demographics of participants.
[DOCX File , 13 KB - mhealth_v9i5e22599_app3.docx]

Multimedia Appendix 4
Table S2: Ownership of smartphones and medical apps.
[DOCX File , 14 KB - mhealth_v9i5e22599_app4.docx]

References

1. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Health Inf Manag 2012;9:1b. [Medline: 22737094]
2. Mobasheri MH, Johnston M, Syed UM, King D, Darzi A. The uses of smartphones and tablet devices in surgery: A systematic review of the literature. Surgery 2015 Nov;158(5):1352-1371. [doi: 10.1016/j.surg.2015.03.029] [Medline: 25964029]
3. Bullock A, Dimond R, Webb K, Lovatt J, Hardyman W, Stacey M. How a mobile app supports the learning and practice of newly qualified doctors in the UK: an intervention study. BMC Med Educ 2015 Apr 08;15:71 [FREE Full text] [doi: 10.1186/s12909-015-0356-8] [Medline: 25889996]
4. Visser BJ, Bouman J. There's a medical app for that. BMJ 2012 Apr 18:e2162. [doi: 10.1136/sbmj.e2162]
5. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 07;1(4):174-181. [doi: 10.1136/bmjinnov-2015-000062]
6. Chen J, Park Y, Putzer G. An Examination of the Components that Increase Acceptance of Smartphones among Healthcare Professionals. Electronic Journal of Health Informatics 2010;5(2):e16 [FREE Full text]

7. El Hadidy TS, Alshafei AE, Mortell AE, Doherty EM. Smartphones in clinical practice: doctors' experience at two Dublin paediatric teaching hospitals. Ir J Med Sci 2018 Aug 23;187(3):565-573. [doi: 10.1007/s11845-017-1713-x] [Medline: 29170970]
8. Ali EE, Chew L, Yap KY. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan 05;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
9. Payne KFB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
10. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: A mixed method study. Health Policy and Technology 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
11. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - what's the evidence? Breast 2014 Oct;23(5):683-689 [FREE Full text] [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
12. Patel RK, Sayers AE, Patrick NL, Hughes K, Armitage J, Hunter IA. A UK perspective on smartphone use amongst doctors within the surgical profession. Ann Med Surg (Lond) 2015 Jun;4(2):107-112 [FREE Full text] [doi: 10.1016/j.amsu.2015.03.004] [Medline: 25905017]
13. Choi JS, Yi B, Park JH, Choi K, Jung J, Park SW, et al. The uses of the smartphone for doctors: an empirical study from samsung medical center. Healthc Inform Res 2011 Jun;17(2):131-138 [FREE Full text] [doi: 10.4258/hir.2011.17.2.131] [Medline: 21886874]
14. Ickin S, Wac K, Fiedler M, Janowski L, Hong J, Dey AK. Factors influencing quality of experience of commonly used mobile applications. IEEE Commun. Mag 2012 Apr;50(4):48-56. [doi: 10.1109/mcom.2012.6178833]
15. Franko OI. Smartphone apps for orthopaedic surgeons. Clin Orthop Relat Res 2011 Jul;469(7):2042-2048 [FREE Full text] [doi: 10.1007/s11999-011-1904-0] [Medline: 21547414]
16. Franko OI, Tirrell TF. Smartphone app use among medical providers in ACGME training programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
17. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
18. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
19. Farnan JM, Snyder Sulmasy L, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627. [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
20. Tran K, Morra D, Lo V, Quan SD, Abrams H, Wu RC. Medical students and personal smartphones in the clinical environment: the impact on confidentiality of personal health information and professionalism. J Med Internet Res 2014 May 22;16(5):e132 [FREE Full text] [doi: 10.2196/jmir.3138] [Medline: 24855046]
21. Plachkinova M, Vo A. Security and Privacy Challenges for Healthcare: Minitrack Overview. 2019 Presented at: 52nd Hawaii International Conference on System Sciences; December 29, 2018; Honolulu, HI p. 4243-4245 URL: https://scholarspace.manoa.hawaii.edu/bitstream/10125/60315/intro-88.pdf [doi: 10.24251/hicss.2019.512]
22. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
23. Blaya JA, Fraser HSF, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251. [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
24. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]

Abbreviations

KHUH: King Hamad University Hospital
mHealth: mobile health
QMH: Queen Mary Hospital

Edited by L Buis; submitted 18.07.20; peer-reviewed by K Wac, A Hidki, A Teles; comments to author 14.09.20; revised version received 08.02.21; accepted 10.03.21; published 10.05.21.

Please cite as:
Nair AA, Afroz S, Ahmed BU, Ahmed UU, Foo CC, Zaidan H, Corbally M
Smartphone Usage Among Doctors in the Clinical Setting in Two Culturally Distinct Countries: Cross-sectional Comparative Study
JMIR Mhealth Uhealth 2021;9(5):e22599
URL: https://mhealth.jmir.org/2021/5/e22599
doi:10.2196/22599
PMID:33970119

Original Paper

Predictors of Walking App Users With Comparison of Current Users, Previous Users, and Informed Nonusers in a Sample of Dutch Adults: Questionnaire Study

Gert-Jan De Bruijn[1], PhD; Joan Martine Dallinga[2], PhD; Marije Deutekom[2], PhD

[1]Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam, Netherlands

[2]Department of Physiotherapy, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

Corresponding Author:
Gert-Jan De Bruijn, PhD
Amsterdam School of Communication Research (ASCoR)
University of Amsterdam
Nieuwe Achtergracht 166
Amsterdam, 1018 WV
Netherlands
Phone: 31 205252636
Email: g.j.debruijn@uva.nl

Abstract

Background: The last decade has seen a substantial increase in the use of mobile health apps and research into the effects of those apps on health and health behaviors. In parallel, research has aimed at identifying population subgroups that are more likely to use those health apps. Current evidence is limited by two issues. First, research has focused on broad health apps, and little is known about app usage for a specific health behavior. Second, research has focused on comparing current users and current nonusers, without considering subgroups of nonusers.

Objective: We aimed to provide profile distributions of current users, previous users, and informed nonusers, and to identify predictor variables relevant for profile classification.

Methods: Data were available from 1683 people who participated in a Dutch walking event in Amsterdam that was held in September 2017. They provided information on demographics, self-reported walking behavior, and walking app usage, as well as items from User Acceptance of Information Technology, in an online survey. Data were analyzed using discriminant function analysis and multinomial logistic regression analysis.

Results: Most participants were current walking app users (899/1683, 53.4%), while fewer participants were informed nonusers (663/1683, 39.4%) and very few were previous walking app users (121/1683, 7.2%). Current walking app users were more likely to report walking at least 5 days per week and for at least 30 minutes per bout (odds ratio [OR] 1.44, 95% CI 1.11-1.85; *P*=.005) and more likely to be overweight (OR 1.72, 95% CI 1.24-2.37; *P*=.001) or obese (OR 1.49, 95% CI 1.08-2.08; *P*=.005) as compared with informed nonusers. Further, current walking app users perceived their walking apps to be less boring, easy to use and retrieve information, and more helpful to achieve their goals. Effect sizes ranged from 0.10 (95% CI 0.08-0.30) to 1.58 (95% CI 1.47-1.70).

Conclusions: The distributions for walking app usage appeared different from the distributions for more general health app usage. Further, the inclusion of two specific subgroups of nonusers (previous users and informed nonusers) provides important information for health practitioners and app developers to stimulate continued walking app usage, including making information in those apps easy to understand and making it easy to obtain information from the apps, as well as preventing apps from becoming boring and difficult to use for goal attainment.

(JMIR Mhealth Uhealth 2021;9(5):e13391) doi:10.2196/13391

KEYWORDS

technology; walking; health; adult; survey; questionnaires

Introduction

Background

The immense popularity of smartphones in the past decade has led to a large number of mobile health apps for those smartphones. Through those apps, people can receive relevant and personalized information and feedback on progress toward public health goals (eg, eating five portions of fruit and vegetables a day, 10,000 steps per day) or self-constructed goals (eg, 10 extra flights of stairs). Mobile health apps tend to focus on physical activity patterns, arguably because the built-in GPS of smartphones allows for an unobtrusive way to monitor those activity patterns. Mirroring the increase in the popularity of mobile health apps, substantial research has investigated the applicability of those health-related mobile apps to influence health behavior in intervention studies or to understand demographic, behavioral, and psychological predictors of mobile health app usage [1-5].

Intervention Effects of Mobile Health Apps

Intervention studies through mobile apps have focused on a wide range of health-related behaviors, such as the treatment of alcohol use disorders [6], weight reduction [3], and physical activity promotion [3,5,7-9]. For physical activity specifically, summary studies have reported positive effects of mobile apps on physical activity [3-5,7]. For instance, Zhao et al [4] found that 17 out of 23 eligible mobile app intervention studies reported positive effects on physical activity behaviors. These findings echo earlier findings from Bort-Roig et al [5], who found that four out of five physical activity mobile app intervention studies increased physical activity, while Fanning et al [7] performed a meta-analysis of 11 mobile app physical activity intervention studies and found a moderate-to-large effect size increase in physical activity for interventions using mobile apps when compared with control conditions. Thus, mobile app interventions appear to be effective when it comes to increasing physical activity behaviors, although some recent systematic reviews have been somewhat less supportive [9-11].

Predictors of Mobile Health App Usage

For predictors of mobile health app usage, most studies have commonly reported that mobile health app users are younger, have higher education, are more involved with or conscious of a healthy lifestyle, and have higher levels of health literacy [1,2,12,13]. For instance, in a population-based survey study among German adults, Ernsting et al [2] found that 20% of smartphone users used health apps on their smartphones and that, compared with nonusers, health app users were younger and more health literate. In addition, health app users were engaging in healthier behaviors. In a similar study on health app usage in South Korea, Cho et al [12] surveyed 765 young adults and found that higher health app usage and stronger perceived efficacy to use health apps were associated with younger age and higher eHealth literacy. Finally, in a population-based survey in over 3500 US adults, Carroll et al [1] found that health app users were younger and had higher education and income. Further, participants who used health apps were more likely to meet physical activity recommendations and have a more positive intention to be physically active. Most study results to date have been mixed regarding the effects of gender on health app usage, with some studies finding gender effects [1] and some not finding these effects [12,13]. Thus, mobile health app users are likely to be younger, have higher education, and be more involved with health-related actions than nonusers.

Current Caveats in Mobile Health App Studies

Despite the informative nature of the current evidence base, there are two important caveats that require further research attention to better understand mobile health app usage. First, most studies on the predictors of mobile health app usage have focused on a simple dichotomy of users versus nonusers. For instance, Carroll et al [1] categorized participants from the 2015 Health Information National Trends Survey in the United States not only along the lines of having a mobile device (ie, mobile or tablet), but also according to whether participants had a health app on their devices. Likewise, Ernsting et al [2] used data from a German population-based sample and compared not only smartphone users and nonusers, but also health app users and nonhealth app users among smartphone users. Such dichotomizations seem to ignore the waxing and waning of a wide range of human health-related actions [14-16], including physical activity [17]. For instance, Conroy et al [17] employed an ecological momentary assessment design among young adults across a 10-week period and found substantial within-person variations in both physical activity intentions and behaviors. These variations presumably also occur for media use [18], but there is remarkably little investigation into specific subgroups of nonusers of mobile health apps. To our knowledge, only one study further divided nonusers into disengaged nonusers (ie, people who have previously used health apps, but no longer use them), informed nonusers (ie, people who have thought about using health apps, but decided not to use them), and unengaged nonusers (ie, people who have never thought about using health apps) [19]. Among the 765 nonusers of apps in this previous study [19], the largest proportions of nonusers were unengaged nonusers (n=301, 39.3%) and informed nonusers (n=214, 28.0%), while a smaller proportion was disengaged users (n=155, 20.2%), with the remaining 95 (12.4%) participants being categorized as participants who decided to start using apps, but were currently not yet using them. Importantly, there were considerable differences between these nonusers in age, with unengaged nonusers being older than most other nonusers. Further, unengaged nonusers reported a healthier eating style than current users, but no differences in eating style emerged between current users and other nonuser profiles. Thus, further differentiation into more specific groups of nonusers seems to be a fruitful research approach to understand the characteristics of subgroups that decide to stop or start using health apps.

A second caveat of the current evidence base is that there is a strong focus on either generic health apps, such as those developed for smoking cessation, dietary change, medication intake, and physical activity [2,12], or apps that focus on a specific behavioral domain, such as general physical activity or dietary intake [20]. However, an investigation of app usage in such a broad behavioral domain tends to neglect various important research findings that have demonstrated poor predictability on broad outcome measures [21] and neglect

theoretical suggestions from behavioral prediction models that tend to require a focus on specific goals and behaviors, such as goal-setting theory [22] and the theory of planned behavior [23]. Moreover, research into the determinants of physical activity behaviors has demonstrated not only that people are more motivated to engage in low-to-moderate physical activities than in vigorous physical activities, but also that motivation-physical activity relationships differ across the levels of physical activity intensity [24,25]. Thus, the current focus in mobile health app usage research on broad behavioral domains (ie, physical activity) may obscure important differences that exists for more specific behaviors in these domains (ie, strenuous exercise versus brisk walking).

Aim

The purpose of this study was threefold. First, we aimed to describe distributions of walking app usage in a sample of Dutch adults, with a specific focus on two important subgroups of nonusers. Because research has indicated that one of the largest subgroups of nonusers is informed nonusers [19], a key focus was to include this subgroup. Likewise, disengaged nonusers represent an important target group, because (in contrast to informed nonusers) they have had previous app experience, but have decided to not continue using the apps. Consequently, they may provide important information on why some people decide to continue using apps and why some do not. Second, we included not only demographic and behavioral correlates as predictors of these subsamples, but also media-use elements of app usage. Media-use elements relate to the experiences that people have when using a specific app [26] or a priori expectations of apps they intend to use [27]. Although media-related experiences [28] are known to be key predictors of continued media usage [29], including health-related media [30] and apps [27], most research on understanding health app usage has focused on either behavioral or demographic correlates, or such factors as health literacy [1,2,19]. Third, we focused on a specific health-related behavior, namely recreational walking. Walking is key to protect against various physical ailments and is associated with various benefits, such as decreased risk factors in type 2 diabetes patients [31] and decreased cardiovascular risk in the general population [32]. Thus, the main aim of this study was to investigate the distributions and predictors of app usage for a specific health behavior, namely walking.

Methods

Participants and Procedures

Data for this study were collected from participants of the Dam tot Dam Wandeltocht (Dam-To-Dam Walk), a walking event in September 2017. This recreational event was organized in Amsterdam, the Netherlands. The inclusion criteria were being 18 years or older and having signed the informed consent form. Three days after the Dam tot Dam Wandeltocht, the target population of the survey (all participants of the Wandeltocht) was invited by email to participate in an online survey that was coordinated by SurveyMonkey [33], using the email address participants provided for their Dam-To-Dam Walk registration. No other announcements or advertisements were made to advertise the study. The email contained a link to a closed online survey, which started with a brief introduction of the study. This introduction provided information about the aim of the study, the right of the participants to quit at any time, and the criteria to participate in the study. After reading the introduction, participants who volunteered to participate in the study provided informed consent. The survey took approximately 15 to 20 minutes to complete, and participants received no reward for survey completion. The timeframe of the survey was September 15, 2017, until October 3, 2017. One week after the first invitation, a reminder email was sent to people who had not responded yet. One week after that, the survey was closed. Participants' IP addresses were registered to prevent people from completing multiple surveys. No other personal information was collected, and IP address data were anonymized. For this study, ethical approval was not obligatory.

Measures

A nonvalidated survey was used to collect data on study variables, which were partly based on previous studies [13,34,35]. Depending on walking app usage status, the survey was either six (current or previous user) or three (informed nonuser) pages long. The number of questions per survey page ranged from one to three, whereas the number of response items per question ranged from one to eight. Most questions did not have a "rather not say" response option, and participants were able to move between survey pages to modify previous answers. The questions and items were not randomized. Two independent researchers outside the author team tested the survey regarding usability and functionality. Data completeness was checked by the first author.

Educational attainment was assessed by presenting participants with a list of commonly completed educational levels in the Netherlands (eg, elementary school, lower vocational education, and university degree). Participants were also able to write down their educational attainment if this list did not include their attained education. The definition from Statistics Netherlands was used to categorize participants as having either low, middle, or high education [36]. Low education was defined as educational attainment ranging from the completion of primary school to the completion of junior classes in secondary education, middle education was defined as the completion of secondary education and middle vocational training, and high education was defined as the completion of higher vocational training, and bachelor and master programs. BMI was calculated from self-reported height and weight following World Health Organization guidelines [37]. Participants were categorized into normal weight, overweight, or obese groups based on those guidelines.

Self-reported walking behavior was assessed by having participants indicate how often they had engaged in recreational walking (1, less than once per month; 2, one to three times per month; 3, once per week; 4, twice per week; 5, three times per week; 6, four or five times per week; 7, more than five times per week), as well as the average walking duration of each bout (1, 0 to 30 minutes; 2, 30 minutes to 1 hour; 3, 1 hour to 90 minutes; 4, 90 minutes to 2 hours; 5, 2 hours to 2.5 hours; 6, 2.5 hours to 3 hours; 7, longer than 3 hours). Participants who

reported engaging in recreational walking at least 5 days per week and longer than at least 30 minutes per bout were categorized as meeting the public health guidelines for moderate physical activity [38].

Walking app usage was assessed by asking participants to tick one of the following four options: (1) I currently use a walking app; (2) I do not currently use a walking app, but I have used one previously; (3) I know of walking apps, but I do not use them; and (4) I do not know of walking apps and I do not use them. Those who used a walking app were further asked to indicate the use duration (1, less than 3 months; 2, 3 to 6 months; 3, 7 to 12 months; 4, 1 to 2 years; 5, 2 to 3 years; 6, 3 to 4 years; 7, 4 to 5 years; 8, longer than 5 years) and frequency (1, all my walking bouts; 2, most of my walking bouts; 3, half of my walking bouts; 4, several of my walking bouts). For frequency, participants were classified as either "always app users" or "fewer than always app users." For duration, participants were classified into three categories (1, duration of 6 months or less; 2, duration between 6 and 12 months; 3, duration longer than 1 year).

App usage experiences were assessed with five items derived from the "User Acceptance of Information Technology" model [34,35]. Participants indicated their agreement (−2, totally disagree; +2, totally agree) with the following items: (1) I find it easy to understand how the app works; (2) It takes a lot of time to understand the app; (3) I find it easy to get information from the app; (4) I find it boring to use the app; and (5) It is easy to achieve my goal with the app. Previous and current users were asked to consider the walking app they used most often, and informed nonusers were asked to consider walking apps in general.

Analysis Plan

Distributions of Walking App Users and Categorical Predictor Variables of Walking App Usage

To investigate univariate distributions of the categorical predictors (gender, weight status, educational attainment, meeting physical activity recommendations, walking app frequency, and walking app duration) across the three app groups, chi-square tests were performed. To investigate the multivariate effect of these same categorical predictors, multinomial logistic regression analysis was performed to calculate the odds of being (1) a current user (reference category), (2) a previous user, and (3) an informed nonuser from gender, education, weight status, and walking behavior. We also investigated the odds of being a current user versus a nonuser, where previous users and noninformed users were collapsed and grouped. Recommended cutoff points for effect sizes for odds ratios (ORs) were followed [39].

Continuous Predictor Variables of Walking App Usage

To predict profile classification from continuous predictors (app experiences, age, and BMI), discriminant function analysis (DFA) was performed. DFA involves a multivariate test to predict group membership based on linear combinations of continuous predictor variables. DFA not only presents the proportion of correctly classified participants, but also reports the correlations of each predictor variable for the discriminant functions. These correlations can be used to interpret the relevance of predictor variables for group membership, with the first function containing the most discriminating predictors. The DFA was followed up by a multivariate analysis of variance, including post-hoc tests using Bonferroni correction and calculation of Cohen *d* [40] to assess the effect sizes of the mean score differences.

Results

Participant Characteristics

Figure 1 presents a flowchart from study invitation to data analysis, with 3435 participants invited to participate. Of those invited, 559 (16.2%) participants declined to participate, 152 (4.4%) did not report any data after accepting the invite, and 465 (13.5%) did not know of mobile walking apps and did not use them. These latter participants were excluded for the purpose of this study. During the remainder of the survey, a further 576 of 2259 (25.5%) participants did not provide information on app use experiences (184/2259, 8.1%) or on demographics or BMI (392/2259, 17.4%). Therefore, the final sample for the main analyses included 1683 participants.

Figure 1. Study flowchart.

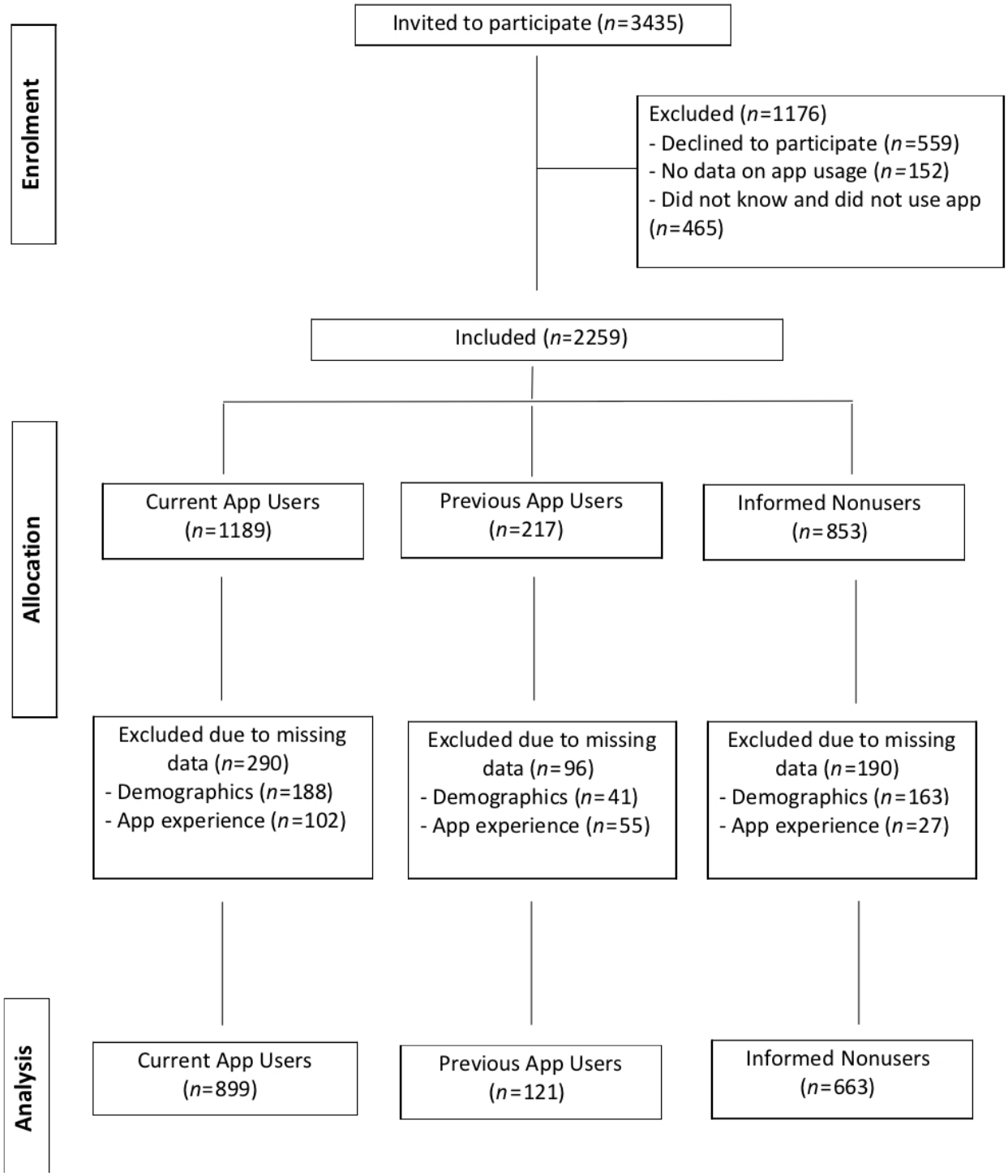


Table 1 presents the characteristics of the total sample and the three profiles. Of the 1683 participants, most were female (1250/1683, 74.3%) and had a middle (758/1683, 45.0%) or high (645/1683, 38.3%) educational level. The mean age was 51.0 years (SD 11.6), and the mean BMI was 26.0 kg/m^2 (SD 4.5). Most participants had a normal weight (772/1683, 45.9%) or were overweight (679/1683, 40.3%), and a small proportion (232/1683, 13.8%) was obese. A little over a fifth of the sample (358/1683, 21.3%) reported walking at least 3 days per week for a minimum of 30 minutes per session. Over half (899/1683, 53.4%) reported currently using an app for walking. Little under 10% (121/1683, 7.2%) reported having previously used an app for walking, and little over a third (663/1683, 39.4%) reported knowing about walking apps but not using them. Among the current and previous users (1020/1683, 60.6%), 558 (54.7%) reported always using a walking app when walking, and nearly two-thirds of users (662/1020, 64.9%) had been using an app for longer than a year.

Table 1. Distributions of categorical variables for the total group and for each of the profiles.

Variable	Total (n=1683), n (%)	Current users (n=899), n (%)	Previous users (n=121), n (%)	Informed nonusers (n=663), n (%)	χ^2 (*df*)	*P* value
Gender					2.1 (2)	.35
Female	1250 (74.3)	652 (72.5)	96 (79.3)	502 (75.7)		
Male	433 (25.7)	247 (27.5)	25 (20.7)	161 (24.3)		
Education					1.4 (4)	.84
Low	280 (16.6)	148 (16.5)	22 (18.2)	110 (16.6)		
Middle	758 (45.0)	398 (44.3)	57 (47.1)	303 (45.7)		
High	645 (38.3)	353 (39.2)	42 (34.7)	250 (37.7)		
Weight status					12.4 (4)	.02
Normal	772 (45.9)	395 (43.9)	46 (38.0)	331 (49.9)		
Overweight	679 (40.3)	363 (40.4)	54 (44.6)	262 (39.5)		
Obese	232 (13.8)	141 (15.7)	21 (17.4)	70 (10.6)		
Walking behavior[a]					6.3 (2)	.04
Meets recommendations	358 (21.3)	208 (23.1)	34 (28.1)	116 (17.5)		
Does not meet recommendations	1325 (78.7)	691 (76.9)	87 (71.9)	547 (82.5)		
Walking app use frequency[b]					55.2 (1)	<.001
Always	559 (54.6)	530 (59.0)	28 (23.1)	N/A[c]		
Less than always	464 (45.4)	369 (41.0)	93 (76.9)	N/A		
Walking app use duration[b]					40.5 (2)	<.001
6 months or less	241 (23.6)	184 (20.5)	55 (45.5)	N/A		
6-12 months	119 (11.6)	103 (11.5)	16 (13.2)	N/A		
More than 1 year	663 (64.8)	612 (68.1)	50 (41.3)	N/A		

[a]Defined as walking at least 5 days per week and at least 30 minutes per day.

[b]Assessed for current and previous users only.

[c]N/A: not applicable.

Categorical Predictor Variables of Walking App Usage

There were significant differences in the distributions of the categorical variables over the three profiles for weight status, walking behavior, and (for current and previous users only) walking app frequency and duration. Specifically, lower proportions of obese participants were found among informed nonusers (70/663, 11%) than among previous (21/121, 17%) or current users (141/899, 15.7%). Likewise, there were lower proportions of participants meeting the activity recommendations among informed nonusers (547/663, 82.5%) than among previous (87/121, 72%) or current users (691/899, 76.9%). Finally, there were higher proportions of people who always used walking apps during their walking routine among current users (530/899, 59.0%) than among previous users (28/121, 23%).

The multinomial logistic regression revealed a significant final model (−2 log likelihood=276.96, χ^2_{12}=30.79, *P*=.002). Meeting the recommendations for physical activity (−2 log likelihood=288.27, χ^2_2=11.31 *P*=.004) and weight status (−2 log likelihood=291.84, χ^2_4=14.88, *P*=.005) were significant contributors to this model, whereas gender (*P*=.13) and educational attainment (*P*=.872) were not significant contributors. Follow-up tests showed that there were no relevant differences between current users and previous users. As compared to informed nonusers, current users were more likely to report having walked at least 5 days per week and for at least 30 minutes per bout (OR 1.44, 95% CI 1.11-1.85; *P*=.005) and more likely to be overweight (OR 1.72, 95% CI 1.24-2.37; *P*=.001) or obese (OR 1.49, 95% CI 1.08-2.08; *P*=.005). Explained variance was low (Nagelkerke R^2=0.02).

The binary logistic regression analysis, where previous users and informed nonusers were collapsed and grouped as nonusers and compared with current users, revealed a similar pattern, with current users being more likely to report having walked at least 5 days per week and for at least 30 minutes per bout (OR 1.29, 95% CI 1.02-1.63; *P*=.04). Nonusers were more likely to have a normal weight as compared to users (OR 1.18, 95% CI 1.02-1.36; *P*=.02). Explained variance was low (Nagelkerke R^2=0.01).

Continuous Predictor Variables of Walking App Usage

There were two significant discriminant functions (function 1: χ^2_{14}=1033.22, *P*<.001, canonical correlation=0.68, Wilk λ=0.54; function 2: χ^2_6=15.03, *P*=.02, canonical correlation=0.09, Wilk λ=0.99). These functions correctly classified 71.5% of cases. Table 2 presents the standardized discriminant coefficients for these functions, the mean scores and standard deviations for each of the predictor variables across the three groups, and the post-hoc comparisons and Cohen *d* for these differences. For the first discriminant function, all app usage coefficients were larger than 0.30, with the strongest predictors being "app boredom" (*r*=0.87), "goal attainment" (*r*=−0.63), "app info retrieval" (*r*=−0.44), and "ease of app usage" (*r*=0.33). For the second discriminant function, age (*r*=0.54) and BMI (*r*=−0.78) were the strongest predictors.

Table 2. Discriminant function analysis.

Predictors	F1[a]	F2[b]	1. Current users, mean (SD)[c]	2. Previous users, mean (SD)[c]	3. Nonusers, mean (SD)[c]	$F_{2, 1680}$	*P* value	η^2	Post-hoc
Easy to understand the app	−0.33	0.33	1.0 (0.8)	0.7 (1.0)	0.4 (0.9)	78.3	<.001	0.09	1>2, *P*=.01 1>3, *P*<.001 2>3, *P*<.001
Takes time to understand the app	0.52	−0.19	−1.2 (0.8)	−0.7 (0.9)	−0.4 (0.9)	189.6	<.001	0.18	1>2, *P*<.001 1>3, *P*<.001 2>3, *P*<.001
Easy to retrieve info from the app	−0.44	0.33	0.9 (0.8)	0.6 (0.9)	0.2 (0.8)	135.3	<.001	0.14	1>2, *P*<.001 1>3, *P*<.001 2>3, *P*<.001
Boring to use the app	0.87	0.23	−1.2 (0.8)	−0.4 (0.9)	0.0 (0.7)	530.3	<.001	0.39	1>2, *P*<.001 1>3, *P*<.001 2>3, *P*<.001
Easy to achieve my goal with the app	−0.63	0.04	0.7 (0.8)	0.1 (0.9)	−0.3 (0.7)	275.7	<.001	0.24	1>2, *P*<.001 1>3, *P*<.001 2>3, *P*<.001
BMI	−0.08	0.54	26.2 (4.5)	26.7 (4.9)	25.5 (4.4)	6.4	.002	0.01	1=2, *P*=.95 1>3, *P*=.005 2>3, *P*=.03
Age	0.13	−0.78	49.8 (11.3)	48.5 (11.9)	52.9 (11.7)	16.9	<.001	0.02	1=2, *P*=.71 1>3, *P*<.001 2>3, *P*<.001

[a]F1 reflects correlation with discriminant function 1.

[b]F2 reflects correlation with discriminant function 2.

[c]Scores for app usage items range from −2 (totally disagree) to +2 (totally agree).

There was a multivariate main effect of app use status ($F_{14,3348}$=86.28, Wilk λ=0.54, *P*<.001, partial η^2=0.27). The univariate tests revealed significant differences between means for all app experience items, which progressed linearly through the three profiles (Table 2). In comparison with previous users and informed nonusers, current app users believed that apps were easy to understand and retrieve information from; helped achieve their goals; were less boring to use; and required less time to understand. The same pattern of differences was found when comparing previous users with informed nonusers. In addition, current users and previous users had higher BMI and were younger as compared with informed nonusers.

Discussion

Distributions of Profiles

The results showed that little over only 5% of participants could be classified as previous users, while a much larger proportion could be classified as informed nonusers and the largest part of the study sample could be classified as current walking app users. These distributions are in line with earlier distributions regarding fitness app usage in German adults [19], which demonstrated that current users and informed nonusers were the most prevalent profiles. Interestingly, these distributions do not completely reflect the distributions found for nutrition apps in that same sample. Only a small proportion was classified as current nutrition app users, while a somewhat larger proportion was classified as previous nutrition app users [19]. The reasons for these discrepant proportions across nutrition and fitness or walking apps are unclear, but they may be explained by the fact

that walking or fitness apps more unobtrusively collect, monitor, and report information regarding either walking (eg, distance and number of steps) or fitness performance (heart rate and heart rate variability), without the need for any input from the user to collect and present this information. In contrast, nutrition apps can mostly only collect and monitor nutrition intake via input from the user, such as textual input of items consumed. Such active involvement in the collection and monitoring of nutrition-related information may put off users from keeping engaged with the nutrition app.

It should also be noted that this study showed higher proportions of current app usage than those reported in other studies, where generally around one-fifth to one-quarter of the study population used apps [1,2]. Although this may be due to the fact that participants who did not use walking apps or did not know about walking apps were excluded, it may reflect increased interest in walking specifically or physical activity in general among those participating in walking events. Indeed, enhanced interest in a subject is known to stimulate learning and information search on that subject [41], both of which can be achieved through health apps [27]. It should also be noted that another recent study in the Dutch population on physical activity app usage found a relatively high proportion of app use [42], which could also indicate that the Dutch population may, in general, be more likely to use physical activity apps.

Demographic and Behavioral Variables as Predictors of Profiles

The results demonstrated that current and previous walking app users not only were much younger than informed nonusers, but also had much higher BMI than informed nonusers. Furthermore, current users were more likely to have an unhealthy weight status (either being overweight or obese) than informed nonusers. These latter findings blend in with previous research in Dutch adults demonstrating that physical activity app users are more likely to report weight loss than nonusers [13]. They also blend in with previous intervention research demonstrating positive effects of app-based technologies on weight and waist circumference [43,44] and correlational research showing that app users are more likely to report more chronic conditions [2] and weight loss [1]. The exact role of walking app usage in weight loss is unclear, but it may suggest that people with an unhealthy BMI may use walking apps more often to obtain the benefits of walking apps. Future research is needed to understand which exact functions in walking apps are used by people with an unhealthy BMI to support them in their weight loss attempts. For instance, walking apps could be used to monitor physical activity levels, set personal and adaptable goals, or monitor progress toward one's walking or weight goal, or could provide an opportunity to receive social support on weight loss through their use.

It should also be noted that this study did not find differences in gender and educational status between the three profiles. This contrasts previous work on mobile health app usage, where occasionally males and more often individuals with lower education [1] or individuals with lower healthy literacy [2] have been found to be less likely to use health apps. However, these earlier studies have collapsed various health-related activities, including smoking control, blood pressure control, medication intake, physical activity, and diet, into their assessments of health app usage [1,2], which prohibits a more fine-grained view of the relation between health behaviors, app usage, and demographics. For instance, there are clear gender differences in adherence to specific health behaviors (eg, walking and vegetable consumption) [45,46], yet these behaviors are often collapsed in studies on health app usage. For instance, Carrol et al [1] investigated app usage for "health-related reasons" and linked differences in app usage to differences in the intention to increase (1) fruit intake, (2) vegetable intake, and (3) physical activity and the intention to decrease (1) soda consumption and (2) weight. Although their findings demonstrated that males were more likely to use these health apps, it is unclear for which specific behaviors these apps were used by those males. Our findings indicate that focus on more specific behaviors in health app usage studies is prudent to better understand the interplay between health behaviors, app usage, and demographics.

When looking at the behavioral predictors, the findings for physical activity levels were mostly in line with previous work, that is, current walking app users were 40% more likely to walk at least 5 days per week and at least 30 minutes per session than informed nonusers. Previous work on understanding usage patterns of both general health apps [1,2] and more behavior-specific apps [47] commonly demonstrated healthier lifestyle patterns for app users. When comparing current users and previous users only, there were some interesting findings regarding previous use duration and frequency. Specifically, among current users, two-thirds reported having used a walking app for more than a year. In contrast, among previous users, this proportion was little over one-third. Perhaps more interesting was the substantial difference in app use frequency between current and previous users. Almost 60% of current users reported always using a walking app during a walking session, whereas less than a quarter of previous users reported always using a walking app. This could potentially indicate a relevant opportunity to enhance walking app user status by stimulating individuals to use their walking app during each and every walking session. Habit formation work could be an interesting area, and future research could consider the inclusion of survey-based habit measures to understand how and when one's walking app usage becomes habitual [48].

App Experience Variables as Predictors of Profiles

The strongest app use experience predictors of profiles were boredom and goal attainment. Participants who were classified as current users perceived the apps to be more helpful in achieving their walking goal and to be less boring to use compared with nonusers, with previous users also perceiving the apps to be more helpful for goal attainment and less boring compared with informed nonusers. These differences were of a large effect size and are particularly informative for walking app design features. Walking app developers should pay particularly close attention to avoid boring elements in their apps and to include aspects that foster walking goal attainment [49]. For instance, boredom has been defined as a state of "lack of stimulation" [50] and as the "aversive experience of wanting but being unable to engage in satisfying activity" [51], which may be alleviated by creating stimulating, engaging, and

enjoyable tasks and environments [52]. Likewise, goals are more likely to be achieved when the set goals are both specific and relatively difficult to achieve [22,53]. Thus, walking app developers should consider creating engaging app environments that foster goal attainment. The results also showed significant differences across the three profiles for how easy it was to understand how to use an app, how long it took for people to understand how an app works, and the ease with which information could be derived from an app, with the highest mean scores for current users and the lowest scores for informed nonusers. There is some evidence that this lack of ease could be related to an overload of cognitive systems by the way the information is presented in competing modalities [54]. This overload can potentially be subverted by presenting information in such a way that textual and visual formats enhance, rather than interfere with, information processing [55]. Thus, a clear understanding of how people process information in mobile apps would be needed to enhance ease of use of health apps.

Limitations

A few limitations need to be mentioned. First, the correlational and cross-sectional design of this study prohibits drawing conclusions about causality between walking app usage, walking, and weight status. A more thorough understanding of the role of walking apps (and health apps in general), walking, and weight status would preferably require longitudinal studies whereby the dynamics of walking, app usage, and weight status are investigated in, for instance, cross-lagged panel designs. Second, the participants were recruited from a public walking event. It may be that the included participants were not an accurate reflection of the general Dutch population in terms of motivation and interest in physical activity and related apps, even though the proportion of participants meeting physical activity recommendations in this study was lower than in other studies on mobile apps and physical activity [1], including a population-based sample [2]. Indeed, when comparing the proportion of app users in this study and in the aforementioned studies [1,2], our study had a much higher proportion of app users. Therefore, caution is needed to generalize these findings. Third, the logistic regression models had very low explanatory value, which is likely related to the absence of additional predictor variables, such as health literacy [2]. Fourth, not all instruments used in this study were validated, and they were mostly derived from more general frequency/duration instruments. Finally, there was an overrepresentation of females in the study sample, which makes generalization to the population at large problematic. It may have been the reason why no gender effects were found on walking app usage.

Conclusions

Despite these limitations, this study demonstrates a few important insights that should help health practitioners to promote walking behavior, as well as help app developers to focus on important media-related technologies that foster walking app usage. It also demonstrates the need to study the predictors of behavior-specific apps, such as walking apps, rather than broader behavioral categories to better understand not only the demographic profiles of people who use those behavior-specific apps, but also how people evaluate those specific apps.

Acknowledgments

G-JDB was supported by the Netherlands Cardiovascular Research Initiative, an initiative with the support of the Dutch Heart Foundation (CVON2016-04) and ZonMw (531003001) in the context of the Supreme Nudge project.

Conflicts of Interest

None declared.

References

1. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
2. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
3. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
4. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
5. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
6. Quanbeck A, Chih M, Isham A, Gustafson D. Mobile Delivery of Treatment for Alcohol Use Disorders: A Review of the Literature. Alcohol Res 2014;36(1):111-122 [FREE Full text] [Medline: 26259005]
7. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]

8. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
9. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
10. Hakala S, Rintala A, Immonen J, Karvanen J, Heinonen A, Sjögren T. Effectiveness of technology-based distance interventions promoting physical activity: Systematic review, meta-analysis and meta-regression. J Rehabil Med 2017 Jan 31;49(2):97-105 [FREE Full text] [doi: 10.2340/16501977-2195] [Medline: 28112356]
11. 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]
12. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 09;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
13. Dallinga JM, Mennes M, Alpay L, Bijwaard H, Baart de la Faille-Deutekom M. App use, physical activity and healthy lifestyle: a cross sectional study. BMC Public Health 2015 Aug 28;15:833-303 [FREE Full text] [doi: 10.1186/s12889-015-2165-8] [Medline: 26316060]
14. Van Zundert RM, Ferguson SG, Shiffman S, Engels R. Dynamic effects of craving and negative affect on adolescent smoking relapse. Health Psychol 2012 Mar;31(2):226-234. [doi: 10.1037/a0025204] [Medline: 21942749]
15. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [doi: 10.1097/01.mlr.0000114908.90348.f9] [Medline: 15076819]
16. Payne N, Jones F, Harris PR. A daily diary investigation of the impact of work stress on exercise intention realisation: can planning overcome the disruptive influence of work? Psychol Health 2010 Jan;25(1):111-129. [doi: 10.1080/08870440903337622] [Medline: 20391210]
17. Conroy D, Elavsky S, Hyde A, Doerksen S. The dynamic nature of physical activity intentions: a within-person perspective on intention-behavior coupling. J Sport Exerc Psychol 2011 Dec;33(6):807-827 [FREE Full text] [doi: 10.1123/jsep.33.6.807] [Medline: 22262706]
18. Hsu MH, Chang CM, Lin HC, Lin YW. Determinants of continued use of social media: the perspectives of uses and gratifications theory and perceived interactivity. Inf Res 2015;20(2):1-14 [FREE Full text]
19. König LM, Sproesser G, Schupp HT, Renner B. Describing the Process of Adopting Nutrition and Fitness Apps: Behavior Stage Model Approach. JMIR Mhealth Uhealth 2018 Mar 13;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
20. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
21. Siegel JT, Navarro MA, Tan CN, Hyde MK. Attitude-behavior consistency, the principle of compatibility, and organ donation: A classic innovation. Health Psychol 2014 Sep;33(9):1084-1091. [doi: 10.1037/hea0000062] [Medline: 24707844]
22. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066x.57.9.705]
23. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
24. Rhodes RE, de Bruijn G. Automatic and motivational correlates of physical activity: does intensity moderate the relationship? Behav Med 2010;36(2):44-52. [doi: 10.1080/08964281003774901] [Medline: 20497942]
25. Koeneman MA, Verheijden MW, Chinapaw MJM, Hopman-Rock M. Determinants of physical activity and exercise in healthy older adults: a systematic review. Int J Behav Nutr Phys Act 2011 Dec 28;8(2):142-152 [FREE Full text] [doi: 10.1186/1479-5868-8-142] [Medline: 22204444]
26. Joo J, Sang Y. Exploring Koreans' smartphone usage: An integrated model of the technology acceptance model and uses and gratifications theory. Computers in Human Behavior 2013 Nov;29(6):2512-2518. [doi: 10.1016/j.chb.2013.06.002]
27. Lee HE, Cho J. What Motivates Users to Continue Using Diet and Fitness Apps? Application of the Uses and Gratifications Approach. Health Commun 2017 Dec;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
28. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth Intervention Using a Smartphone App to Increase Walking Behavior in Young Adults: A Pilot Study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
29. Katz E, Blumler J, Gurevitch M. Uses and Gratifications Research. Public Opinion Quarterly 1973;37(4):509-523. [doi: 10.1086/268109]
30. Alpert JM, Krist AH, Aycock RA, Kreps GL. Designing User-Centric Patient Portals: Clinician and Patients' Uses and Gratifications. Telemed J E Health 2017 Mar;23(3):248-253 [FREE Full text] [doi: 10.1089/tmj.2016.0096] [Medline: 27333468]

31. Qiu S, Cai X, Schumann U, Velders M, Sun Z, Steinacker JM. Impact of walking on glycemic control and other cardiovascular risk factors in type 2 diabetes: a meta-analysis. PLoS One 2014;9(10):e109767 [FREE Full text] [doi: 10.1371/journal.pone.0109767] [Medline: 25329391]
32. Murtagh EM, Nichols L, Mohammed MA, Holder R, Nevill AM, Murphy MH. The effect of walking on risk factors for cardiovascular disease: an updated systematic review and meta-analysis of randomised control trials. Prev Med 2015 Mar;72:34-43. [doi: 10.1016/j.ypmed.2014.12.041] [Medline: 25579505]
33. SurveyMonkey. URL: https://nl.surveymonkey.com/ [accessed 2021-04-15]
34. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
35. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
36. Standard Education Grouping 2016. Edition 2017-2018. Den Haag/Heerlen: Statistics Netherlands; 2018.
37. Physical status: the use and interpretation of anthropometry: Report of a WHO Expert Committee. World Health Organization. URL: https://www.who.int/childgrowth/publications/physical_status/en/ [accessed 2021-04-14]
38. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
39. Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. Communications in Statistics - Simulation and Computation 2010 Mar 31;39(4):860-864. [doi: 10.1080/03610911003650383]
40. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
41. Silvia PJ. Interest—The Curious Emotion. Curr Dir Psychol Sci 2008 Feb;17(1):57-60. [doi: 10.1111/j.1467-8721.2008.00548.x]
42. Dallinga J, Van Rijn R, Stubbe J, Deutekom M. Injury incidence and risk factors: a cohort study of 706 8-km or 16-km recreational runners. BMJ Open Sport Exerc Med 2019;5(1):e000489 [FREE Full text] [doi: 10.1136/bmjsem-2018-000489] [Medline: 30899549]
43. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
44. Orsama A, Lähteenmäki J, Harno K, Kulju M, Wintergerst E, Schachner H, et al. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: results of a theory-based randomized trial. Diabetes Technol Ther 2013 Aug;15(8):662-669. [doi: 10.1089/dia.2013.0056] [Medline: 23844570]
45. Prättälä R, Berg MA, Puska P. Diminishing or increasing contrasts? Social class variation in Finnish food consumption patterns, 1979-1990. Eur J Clin Nutr 1992 Apr;46(4):279-287. [Medline: 1600925]
46. Pollard TM, Wagnild JM. Gender differences in walking (for leisure, transport and in total) across adult life: a systematic review. BMC Public Health 2017 Apr 20;17(1):341 [FREE Full text] [doi: 10.1186/s12889-017-4253-4] [Medline: 28427376]
47. Bruno M, Wright M, Baker CL, Emir B, Carda E, Clausen M, et al. Mobile App Usage Patterns of Patients Prescribed a Smoking Cessation Medicine: Prospective Observational Study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e97 [FREE Full text] [doi: 10.2196/mhealth.9115] [Medline: 29666043]
48. Gardner B, Abraham C, Lally P, de Bruijn G. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012 Aug 30;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
49. Locke EA, Latham GP. New Directions in Goal-Setting Theory. Curr Dir Psychol Sci 2016 Jun 23;15(5):265-268. [doi: 10.1111/j.1467-8721.2006.00449.x]
50. Fenichel O. On the psychology of boredom. In: Rapaport D, editor. Organization and pathology of thought: Selected sources. New York, NY, USA: Columbia University Press; 1951:349-361.
51. Eastwood JD, Frischen A, Fenske MJ, Smilek D. The Unengaged Mind: Defining Boredom in Terms of Attention. Perspect Psychol Sci 2012 Sep;7(5):482-495. [doi: 10.1177/1745691612456044] [Medline: 26168505]
52. Lyons EJ. Cultivating Engagement and Enjoyment in Exergames Using Feedback, Challenge, and Rewards. Games Health J 2015 Feb;4(1):12-18 [FREE Full text] [doi: 10.1089/g4h.2014.0072] [Medline: 26181675]
53. Epton T, Currie S, Armitage CJ. Unique effects of setting goals on behavior change: Systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]
54. Meppelink CS, Smit EG, Diviani N, Van Weert JCM. Health Literacy and Online Health Information Processing: Unraveling the Underlying Mechanisms. J Health Commun 2016;21(sup2):109-120. [doi: 10.1080/10810730.2016.1193920] [Medline: 27668318]
55. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]

Abbreviations

DFA: discriminant function analysis
OR: odds ratio

Edited by M Focsa; submitted 14.01.19; peer-reviewed by L König, J Bort-Roig, DR Wahl; comments to author 20.02.19; revised version received 19.05.19; accepted 19.06.19; published 12.05.21.

Please cite as:
De Bruijn GJ, Dallinga JM, Deutekom M
Predictors of Walking App Users With Comparison of Current Users, Previous Users, and Informed Nonusers in a Sample of Dutch Adults: Questionnaire Study
JMIR Mhealth Uhealth 2021;9(5):e13391
URL: https://mhealth.jmir.org/2021/5/e13391
doi:10.2196/13391
PMID:33978595

Original Paper

The Effect of Sensor Placement and Number on Physical Activity Recognition and Energy Expenditure Estimation in Older Adults: Validation Study

Anis Davoudi[1], PhD; Mamoun T Mardini[2], PhD; David Nelson[3], MBA; Fahd Albinali[3], PhD; Sanjay Ranka[4], PhD; Parisa Rashidi[1], PhD; Todd M Manini[2], PhD

[1]Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States

[2]Department of Aging and Geriatric Research, University of Florida, Gainesville, FL, United States

[3]Qmedic Medical Alert Systems, Boston, MA, United States

[4]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States

Corresponding Author:
Anis Davoudi, PhD
Department of Biomedical Engineering
University of Florida
M542, Stetson Medical Science Building
1345 Center Dr
Gainesville, FL, 32610
United States
Phone: 1 352 294 5086
Email: anisdavoudi@ufl.edu

Abstract

Background: Research has shown the feasibility of human activity recognition using wearable accelerometer devices. Different studies have used varying numbers and placements for data collection using sensors.

Objective: This study aims to compare accuracy performance between multiple and variable placements of accelerometer devices in categorizing the type of physical activity and corresponding energy expenditure in older adults.

Methods: In total, 93 participants (mean age 72.2 years, SD 7.1) completed a total of 32 activities of daily life in a laboratory setting. Activities were classified as sedentary versus nonsedentary, locomotion versus nonlocomotion, and lifestyle versus nonlifestyle activities (eg, leisure walk vs computer work). A portable metabolic unit was worn during each activity to measure metabolic equivalents (METs). Accelerometers were placed on 5 different body positions: wrist, hip, ankle, upper arm, and thigh. Accelerometer data from each body position and combinations of positions were used to develop random forest models to assess activity category recognition accuracy and MET estimation.

Results: Model performance for both MET estimation and activity category recognition were strengthened with the use of additional accelerometer devices. However, a single accelerometer on the ankle, upper arm, hip, thigh, or wrist had only a 0.03-0.09 MET increase in prediction error compared with wearing all 5 devices. Balanced accuracy showed similar trends with slight decreases in balanced accuracy for the detection of locomotion (balanced accuracy decrease range 0-0.01), sedentary (balanced accuracy decrease range 0.05-0.13), and lifestyle activities (balanced accuracy decrease range 0.04-0.08) compared with all 5 placements. The accuracy of recognizing activity categories increased with additional placements (accuracy decrease range 0.15-0.29). Notably, the hip was the best single body position for MET estimation and activity category recognition.

Conclusions: Additional accelerometer devices slightly enhance activity recognition accuracy and MET estimation in older adults. However, given the extra burden of wearing additional devices, single accelerometers with appropriate placement appear to be sufficient for estimating energy expenditure and activity category recognition in older adults.

(JMIR Mhealth Uhealth 2021;9(5):e23681) doi:10.2196/23681

KEYWORDS

human activity recognition; machine learning; wearable accelerometers; mobile phone

Introduction

Background

Over the past 30 years, accelerometer devices have been widely used for measuring movements, physical activity categories, and energy expenditure [1]. This work has also carried forward into characterizing the activity patterns of patients with chronic diseases such as obesity, cardiovascular disease, schizophrenia, bipolar disorder, and cancer [2-6]. Despite its growing use in both clinical and research settings, the optimal body position for sensor placement that would provide the most accurate activity category recognition and the corresponding estimate of energy expenditure in older adults remains uncertain. For example, previous studies have used various sensor placements on the body, including the wrist [7-9], thigh [10,11], hip [12-14], arm [15,16] or ankle [17,18], or a combination of multiple placements [19,20]. However, such studies have often been conducted on relatively small samples of young and middle-aged adults. There continues to be a gap in knowledge regarding body placement for older adults (>60 years). Such knowledge is important for considering older age as a factor for estimating activity types and energy expenditure.

There is a lack of a comprehensive evaluation that directly compares individual and combinations of accelerometers placed on different body positions. Historically, the hip position was chosen in both research and public settings for tracking steps (ie, steps per day). The hip position is close to the body's center of the mass and provides an acceleration change because of the foot fall action-reaction when ambulating. As such, the hip position offers a convenient and accurate approach for capturing ambulatory activity [21]. The ankle position is also accurate in assessing step counts and other gait-related features [22-25]. Recently, however, the wrist position has become popular for collecting accelerometer data because of the increased prevalence of smartwatches. This is due to their convenience, ability to capture sleep quality, determination of 24-hour activity rhythms, and enhanced compliance [26-30].

Objectives

A systemic evaluation of body placements will help optimize energy expenditure estimation and activity recognition. It would also help resolve controversies related to the balance between the accuracy and convenience of different body placements [31]. Given the paucity of information about the role of accelerometer placement on older adults, we aimed to compare and contrast energy expenditure estimation, individual activity, and activity category recognition with 5 sensor body positions and their combinations during 32 activities that included sedentary, locomotion, and lifestyle categories. We hypothesized that combined data from 5 accelerometer positions on the body would provide optimal energy expenditure estimation, individual activity recognition, and activity category recognition, but this improvement will be incremental compared with a single or combination of body placements.

Methods

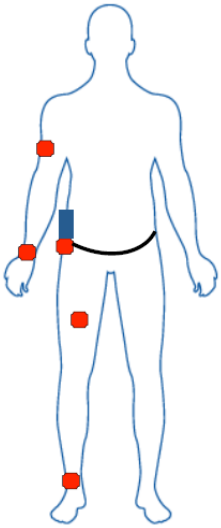
Study Design

This study was approved by the University of Florida Institutional Review Board, and written informed consent was obtained from all participants. The inclusion criteria were designed to optimize safety while ensuring population representation. It included older adults, aged ≥60 years [32], with stratified enrollment for both high and low function according to scores on the standardized Short Physical Performance Battery [33]. The study pre-planned to enroll and complete testing in 90 participants with 30% (27/90) of the participants scoring in the lowest quartile of physical function. Recruitment focused on enrolling community-dwelling adults without significant health issues that could impact the safety of participants. Additional inclusion criteria included willingness to undergo all testing procedures, stable weight for at least 3 months, and ability to understand and speak English. Participants were excluded if they met any of the following criteria: failure to provide informed consent, use of a walker, lower extremity amputation, history of chest pain or severe shortness of breath during physical stress, poststroke syndrome causing ambulatory deficits, and requiring assistance with basic activities of daily living or living in a complete care nursing home. A complete list of the exclusion criteria can be found elsewhere [34].

Accelerometers and Energy Expenditure During Activities

Participants were asked to perform 32 scripted activities listed in Multimedia Appendix 1. These activities were chosen because they are common among most Americans and are consistent with the average time spent in the 2010 American Time Use Survey [35]. Activities were performed for 6 to 8 minutes with 5 to 10 minutes of rest between each activity. Assessments were completed over 4 separate visits. The participants received instructions from the research staff before each activity. Participants wore 5 ActiGraph GT3X triaxial accelerometers [36], one on their ankle, upper arm, hip, thigh, and wrist. All monitors were worn on the right side for the duration of data collection, as shown in Figure 1. Of note, Buchan et al [37] and Dieu et al [38] demonstrated strong agreement between accelerometer data collected on the dominant and nondominant sides. Accelerometers were initialized simultaneously and programmed to collect data at 100 Hz.

Figure 1. Sensor placement on the body.



Participants wore a COSMED K4b2 [39] portable gas analysis system while performing the 32 scripted activities. Before data collection, the oxygen (O_2) and carbon dioxide (CO_2) sensors were calibrated using a gas mixture sample of 16.0% O_2 and 5.0% CO_2 and room air calibration. The turbine flow meter was calibrated using a 3.0-L syringe. A flexible facemask was positioned over the participant's mouth and nose and attached to the flow meter. Oxygen consumption (VO_2; measured in mL min^{-1} kg^{-1}) was measured breath-by-breath, and data were subsequently smoothed with a 30-second running average window. VO_2 data were displayed and manually evaluated to determine when steady-state VO_2 was reached. A steady state was defined as a plateau in VO_2, which typically occurs 2 minutes after the start of the activity. Data were expressed as metabolic equivalents (METs) after dividing the VO_2 values by the traditional standard of 3.5 mL min^{-1} kg^{-1} [40]. A dedicated study smartphone with a custom-built app was synchronized to server time and used to record the start and stop times for each activity (shown in blue in Figure 1). This ensured that time windows could be accurately identified from accelerometer data that was also initialized to server time.

Analysis

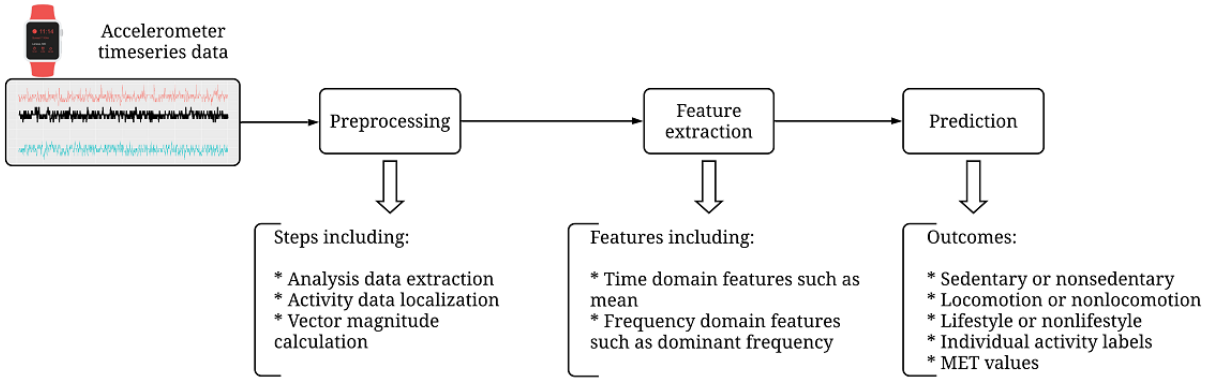
Data were first processed to extract relevant summary features from each contiguous 16-second window. The features described in Table 1 represent both the time and frequency domains [41,42]. These features were included in the analytic models, as illustrated in the analysis flow in Figure 2. There were a total of 31 different wrist, hip, ankle, upper arm, and thigh body position combinations. The analyses compared the performance of single placement and combinations of device placements for estimating METs and for labeling activities as individual and when they were categorized as sedentary, locomotion, or lifestyle (Multimedia Appendix 1). We used random forest as our primary analysis approach, which is a frequently used machine learning algorithm, to recognize human activity from accelerometer data [41-45]. Random forest is an ensemble learning algorithm that builds a large number of decision trees from random sub–data sets of the training data set. The predicted class is determined by aggregating the predicted classes (votes) from the individual decision trees and selecting the majority class in case of classification or by averaging the predicted values in case of regression [46]. This procedure was first performed to evaluate the accuracy of detecting activity categories based on sedentary versus nonsedentary, locomotion versus nonlocomotion, and lifestyle versus nonlifestyle activities as well as to evaluate the accuracy of classifying each of the 32 individual activities against a 3.1% random chance of matching correctly. We used a regression random forest for continuous MET estimation and classification of random forest for activity recognition. To reduce bias, the data were split randomly into development and testing data sets using participant identification numbers. Participants were included in either the development or testing data sets but not both. The development data set was further randomly split into training and validating data sets to tune the model parameters. Nested cross-validation was used; in each outer fold, we kept five-sixths of the participants for model development and one-sixth of the participants for testing. In each inner fold, four-fifths of the participants in the development data set were assigned to the training data set, and one-fifth of the participants were assigned to the validating data set. All model estimates were reported for the testing data sets. In supplementary analyses, a confusion matrix of actual versus predicted activities (32×32 matrix) from the hip and wrist positions, respectively, was generated to help interpret the accuracy and F1 score results. We chose to examine these positions because they are the most used in the literature.

Table 1. Description of features extracted from the raw data.

Feature	Description
Time	
Mean of vector magnitude	Sample mean of the VM[a] in the window
SD of vector magnitude	SD of VM in the window
Mean angle of acceleration relative to vertical on the deice	Sample mean of the angle between x-axis and VM in the window
SD of the angle of acceleration relative to vertical on the device	Sample SD of the angles in the window
Covariance	Covariance of the VM in the window
Skewness	Skewness of the VM in the window
Kurtosis	Kurtosis of the VM in the window
Entropy	Entropy of the VM in the window
Coefficient of variation	SD of VM in the window divided by the mean, multiplied by 100
Corr(x,y)	Correlation between x-axis and y-axis
Corr(y,z)	Correlation between y-axis and z-axis
Corr(x,z)	Correlation between x-axis and z-axis
Frequency	
Percentage of the power of the VM that is in 0.6-2.5 Hz	Sum of moduli corresponding to frequency in this range divided by sum of moduli of all frequencies
Dominant frequency of VM	Frequency corresponding to the largest modulus
Fraction of power in VM at dominant frequency	Modulus of the dominant frequency or sum of moduli at each frequency

[a]VM: vector magnitude.

Figure 2. Analysis flow steps. After accelerometer data were downloaded using the ActiLife (ActiGraph) toolbox, preprocessing steps and feature extraction steps were completed to prepare the data set to be used in prediction models for each task. MET: metabolic equivalent.



Model Evaluation

We calculated the performance metrics of the models by comparing the model-based predicted values with the measured values. For the performance of the individual activity recognition model, we calculated the total accuracy of the model. For activity category recognition, we used the balanced accuracy metric to report model performance because of the class imbalance (ratio of the majority class to minority class being much smaller than 1) across activities. Balanced accuracy is defined as the mean of sensitivity and specificity metrics [47,48]. For MET estimation, we used the predicted and measured values to calculate the root mean square error (RMSE). The results were summarized into 3 major categories: the most accurate combination, the most accurate placement performance, and the most efficient combination. The latter was defined as the fewest number of sensors that provide a similar performance to the most accurate combination, with less than a 10% decrease in performance compared with the most accurate combination. For visualization purposes, the difference in the balanced accuracy of body placement/s compared with the accuracy derived from all 5 sensors was plotted. They were grouped by the number of body placements and ranked to simplify the visual comparisons. To compare across figures, the absolute value of the individual balanced accuracy was also added to the illustration.

Results

The study enrolled 93 older adults (mean age 72.2, SD 7.1 years). The sample was balanced across gender, was mostly non-Hispanic White, and had comorbidities similar to those of the general population. Table 2 presents the descriptive characteristics of the participants. The participants completed 2013 tasks. The median number of tasks completed was 26 out of 32 tasks (Multimedia Appendix 1). Stair ascent had the lowest amount of complete data (n=43) and leisure walk had the most complete data (n=82). The reasons for missing information included not reaching a steady-state metabolic rate, invalid data from one or more monitors, unable to complete the task for at least 4 minutes, missed visits, or provided only partial data because the participant withdrew from the study.

Table 2. Participant characteristics (n=93).

Characteristics	Values
Age (years), mean (SD)	72.17 (7.02)
Female, n (%)	47 (51)
BMI (kg/m^2), mean (SD)	28.18 (4.92)
Race or ethnicity, n (%)	
Non-Hispanic White	83 (89)
Non-Hispanic Black	8 (9)
Non-Hispanic Asian	1 (1)
Hispanic	2 (2)
Education (≥16 years), n (%)	15 (16)
Married or in a relationship, n (%)	52 (56)
Live alone, n (%)	30 (32)
Household income (≥US $15,000), n (%)	66 (71)
Self-rated health (≥good), n (%)	87 (94)
Self-reported conditions, n (%)	
Former or current smoker	37 (40)
Hypertension	45 (48)
Hypercholesterolemia	39 (42)
Diabetes	19 (20)
Chronic pulmonary disease	10 (11)
Heart attack, myocardial infarction	8 (9)
Cancer	27 (29)
Depression	10 (11)
Stroke	4 (4)
Osteoarthritis	11 (12)
Total moderate physical activity (min/week)[a]	93.50
Walking speed (min per second), mean (SD)	
Leisure pace[b]	1.29 (0.26)
Rapid pace[c]	1.41 (0.25)

[a]Data included for 77 participants.

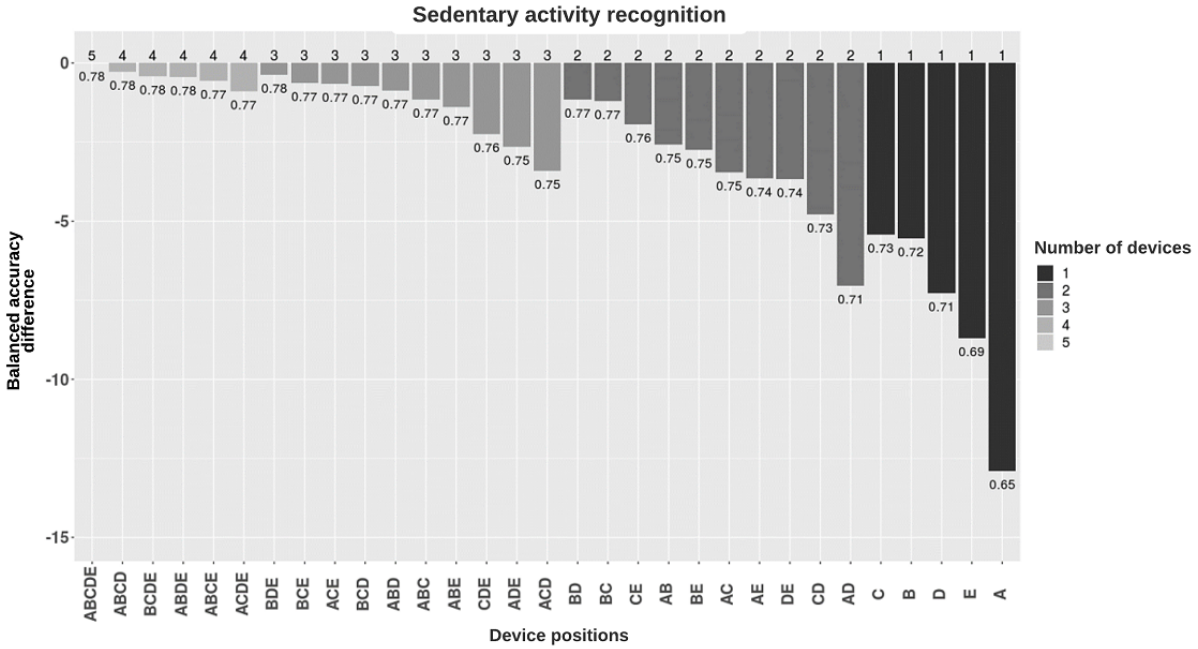
[b]Data included for 91 participants.

[c]Data included for 85 participants.

Models were also tested for categorizing sedentary, locomotion, and lifestyle activities (Figures 3-5). For sedentary behavior recognition, the combination of all accelerometers resulted in the best performance (balanced accuracy 0.78). Hip-worn placement provided the best performance among the single-placement models (balanced accuracy 0.73). The ankle-worn placement resulted in the worst performance (balanced accuracy 0.65). Multimedia Appendices 2 and 3 illustrate confusion matrices of the hip and wrist positions revealing that strength exercise and yoga, both partially done

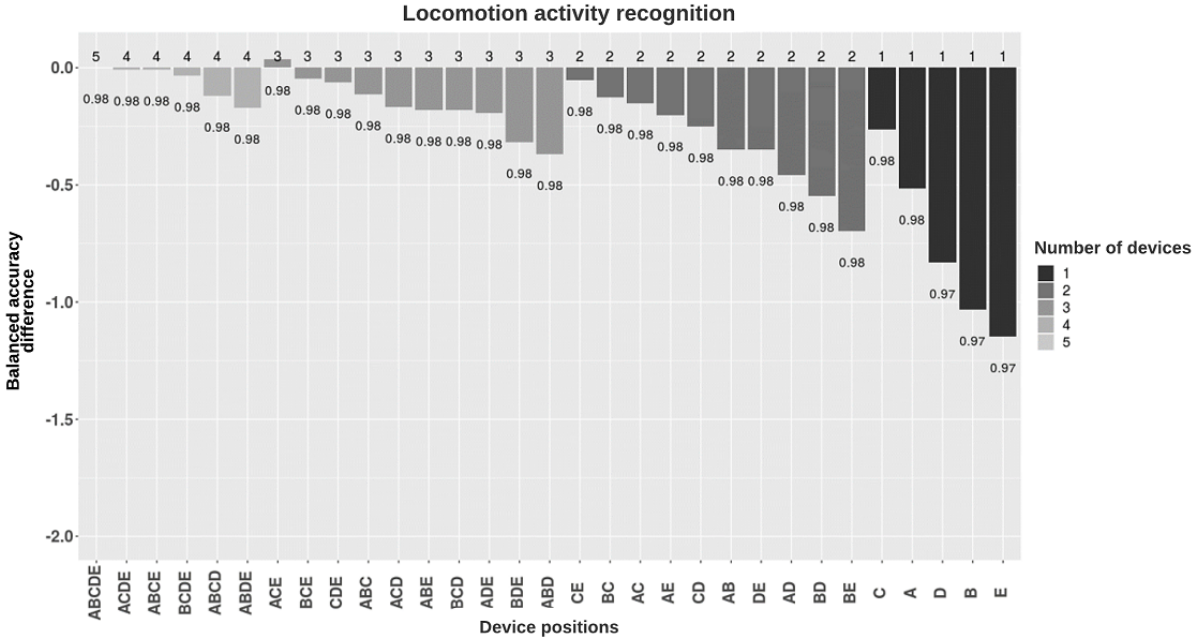
in a sitting position, were mislabeled as being sedentary activities, which caused significant overall misclassification.

Figure 3. Balanced accuracy performance of sedentary activity classification models based on the device placement combinations. Models were grouped by the number of devices used and, in each group, were sorted by decreasing balanced accuracy (rounded). Y-axis shows the difference between the balanced accuracies of the different combinations and the five-placement combination. Numbers in the plot show the balanced accuracies of each placement combination. A: ankle; B: upper arm; C: hip; D: thigh; E: wrist.



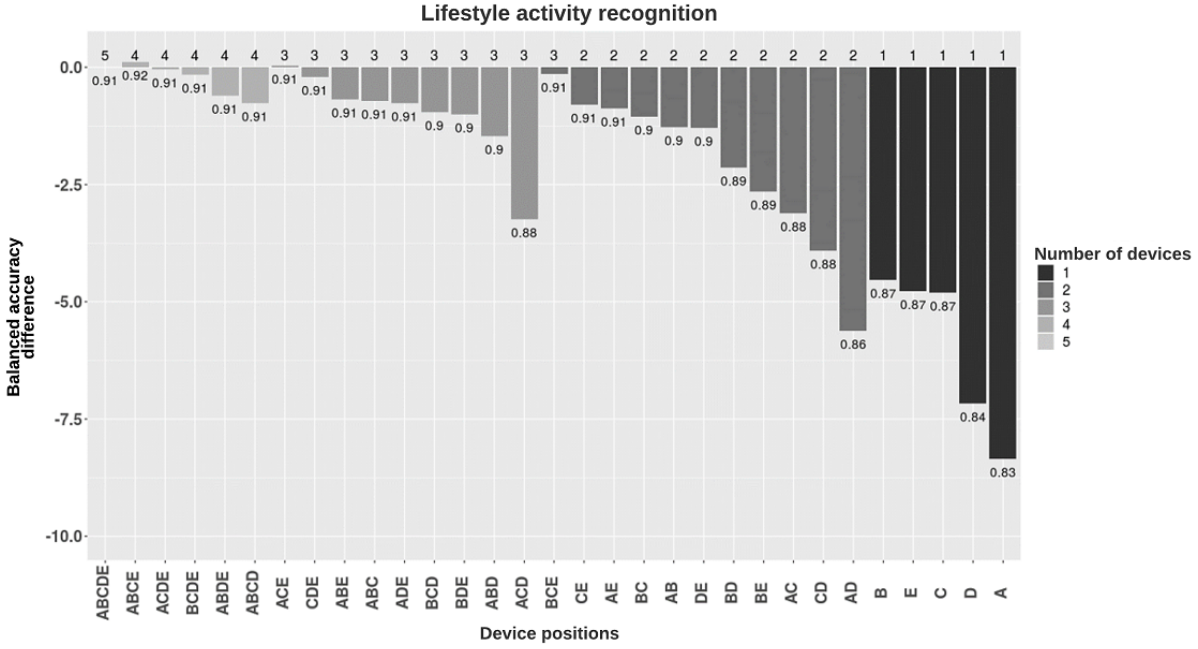
For locomotion activity recognition, the combination of all placements resulted in the best performance (balanced accuracy 0.98). Hip-worn placement provided the best performance among the single-placement models (balanced accuracy 0.98). Classifiers trained separately on data from ankle-worn, wrist-worn, arm-worn, and thigh-worn placement also resulted in high performance (balanced accuracy 0.97-0.98; Figure 4).

Figure 4. Balanced accuracy performance of locomotion activity classification models based on the device placement combinations. Models were grouped by the number of devices used and, in each group, were sorted in decreasing balanced accuracy (rounded). Y-axis shows the difference between the balanced accuracies of the different combinations and the five-placement combination. Numbers in the plot show the balanced accuracies of each placement combination. A: ankle; B: upper arm; C: hip; D: thigh; E: wrist.



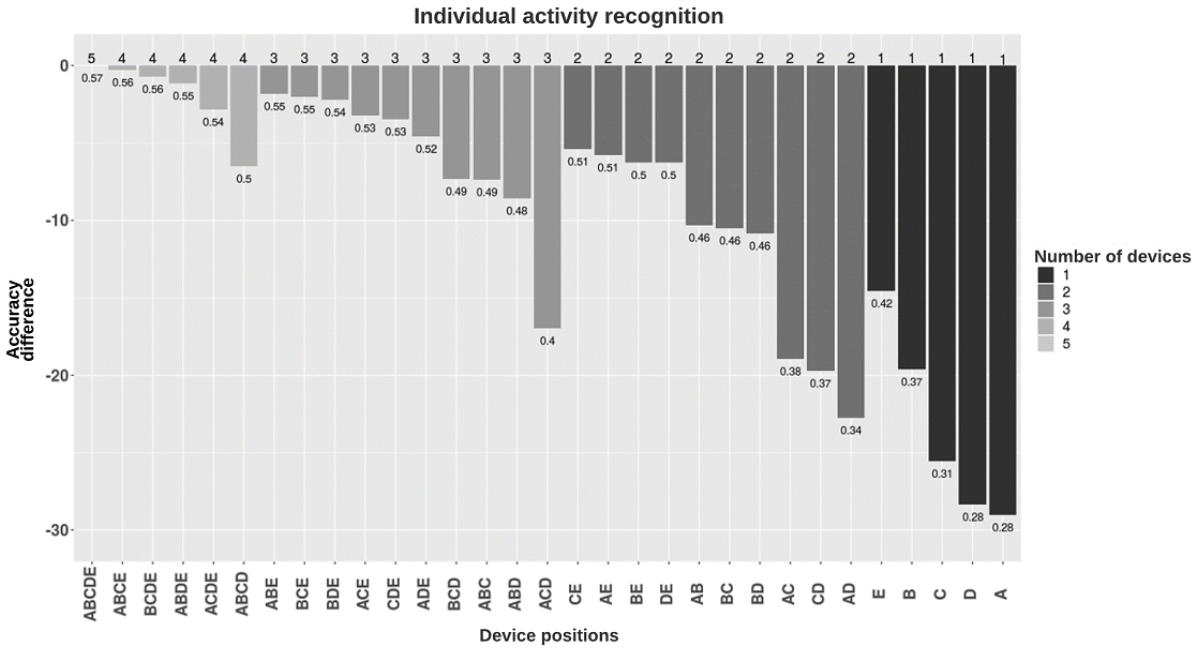
For lifestyle activity recognition, the combination of data from ankle-worn, arm-worn, hip-worn, and wrist-worn placements resulted in the best performance (balanced accuracy 0.92). The combination of data from all placements resulted in high performance (balanced accuracy 0.91). Classifiers trained on data from arm-worn placements, similar to hip-worn and wrist-worn placements, provided the best performance among the single-placement models (balanced accuracy 0.87), whereas ankle-worn placement resulted in the lowest performance (balanced accuracy 0.83; Figure 5).

Figure 5. Balanced accuracy performance of lifestyle activity classification models based on the device placement combinations. Models were grouped by the number of devices used and, in each group, were sorted in decreasing balanced accuracy (rounded). Y-axis shows the difference between the balanced accuracies of the different combinations and the five-placement combination. Numbers in the plot show the balanced accuracies of each placement combination. A: ankle; B: upper arm; C: hip; D: thigh; E: wrist.



The individual activity recognition models with all placements resulted in a relatively low accuracy of 0.57 (Figure 6). Wrist-worn placement provided the best performance among the single-placement models (accuracy 0.42). Classifiers trained separately on data from the ankle-worn placement, similar to thigh-worn placement, resulted in the worst performance (accuracy 0.28; Figure 6).

Figure 6. Accuracy performance of individual activity classification models based on the device placement combinations. Models were grouped by the number of devices used and, in each group, were sorted in decreasing accuracy (rounded). Y-axis shows the difference between the accuracies of the different combinations and the five-placement combination. Numbers in the plot show the accuracy of each placement combination. A: ankle; B: upper arm; C: hip; D: thigh; E: wrist.



Energy expenditure accuracy was evaluated using the MET RMSE of the predicted versus measured values (Figure 7). In general, models trained using the combination of data from all 5 placements resulted in an RMSE of 0.88 METs. Hip-worn and thigh-worn placements provided the lowest RMSE of 0.91 METs among the single body placements. Overall, there was a slight reduction in RMSE when additional accelerometer placement was added to the model.

Figure 7. Root mean square error (RMSE) score performance of met value estimation models based on the device placement combinations. Models were grouped by the number of devices used and, in each group, were sorted in increasing RMSE (rounded). Y-axis shows the difference between the RMSE values of the different combinations and the five-placement combination. A: ankle; B: upper arm; C: hip; D: thigh; E: wrist; MET: metabolic equivalent; RMSE: root mean square error.

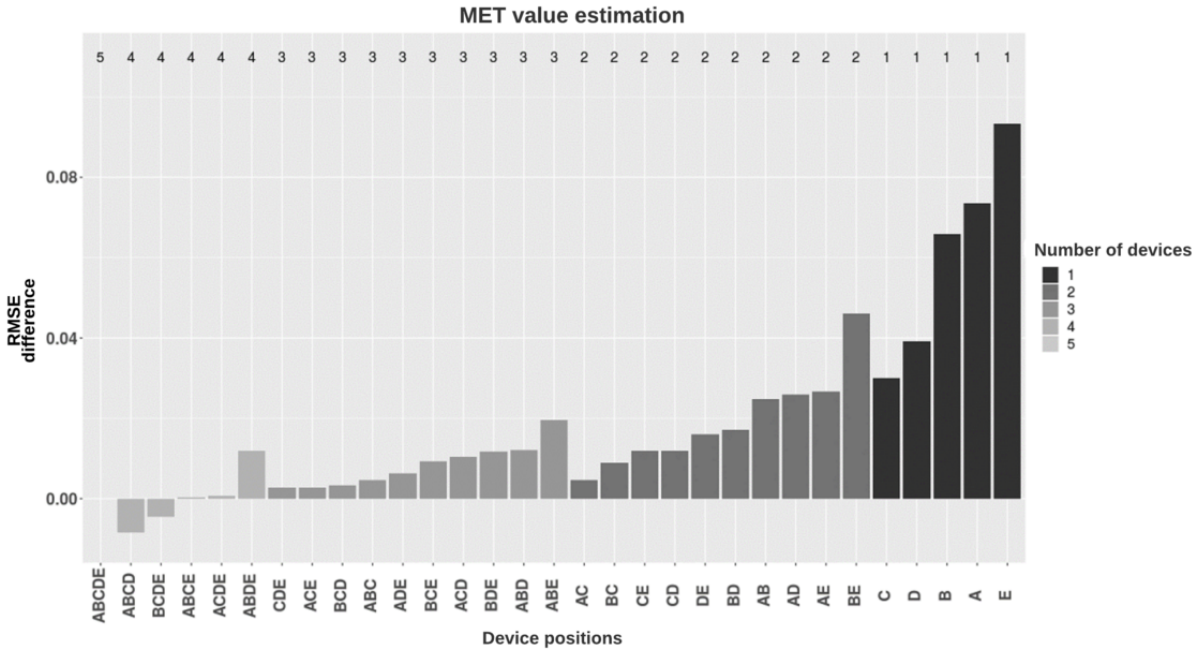


Table 3 summarizes the results according to the positions deemed most accurate, best single placement, and most efficient combination. In general, the most accurate combination contained data from all 5 body positions, but the most accurate placement was often very similar and sometimes better than combinations. The hip and wrist positions appeared to be the most efficient combinations, but models were able to recognize individual activities only with chance probability.

Table 3. Guideline table to determine the needed number and placement of the wearable accelerometer for each task.

Task	Most accurate combination	Most accurate single placement	Most efficient combination[a]
Sedentary activity detection (balanced accuracy)	All 5 placements (0.78)	Hip (0.73)	Hip (0.73)
Locomotion activity detection (balanced accuracy)	All 5 placements (0.98)	Hip (0.98); ankle (0.98)	Hip (0.98)
Lifestyle activity detection (balanced accuracy)	Ankle+upper arm+hip+wrist (0.92)	Upper arm (0.87); wrist (0.87); hip (0.87)	Wrist (0.87)
Individual activity recognition (accuracy)	All 5 placements (0.57)	Wrist (0.42)	Hip+wrist (0.51)
MET value estimation (root mean square error)	Ankle+upper arm+hip+thigh[b] (0.87)	Hip (0.91); thigh (0.91)	Hip+wrist (0.89)

[a]The most efficient combination was defined as the fewest number of sensors that provide a similar performance to the most accurate combination while considering usability. Similar performance was defined as a difference ≤10% of the most accurate combination. We considered the most-to-least usable placements to be wrist>hip>ankle>arm>thigh. Thus, if the performance difference was less than 10%, then the most usable placement was chosen as the most efficient. Best and worst performance refer to best and worst performance according to their balanced accuracy (best: highest balanced accuracy; worst: lowest balanced accuracy).

[b]The performance of the combination with the best performance (0.87) was very close to that of the combination with all 5 placements (0.88).

Discussion

Principal Findings

We compared the performance of activity recognition models based on different combinations of 5 accelerometer placements on 32 activities of daily life. We considered single-sensor and multisensor placement on the wrist, hip, ankle, upper arm, and thigh. Our results show that the models achieved the best performance in the classification of locomotion activities and lifestyle activities (balanced accuracies 0.98 and 0.91 for the all five-sensor combination, respectively), followed by the classification of sedentary activity (balanced accuracy 0.78). The correct labeling of individual activities was low (accuracy

0.57). Interestingly, increasing the number of accelerometer placements had very limited improvement in the classification accuracy of locomotion and lifestyle activities as well as estimating MET values.

There are also noteworthy results from locomotion and sedentary tasks. The accuracy of locomotion activity recognition was similar across all the placements, and only minor differences were found between the combinations (approximately 1%). It is worth mentioning that the wrist-worn accelerometer had relatively lower performance, which is potentially due to the locomotor-like hand movements observed in other nonlocomotor tasks (eg, washing windows and yard work). Nonetheless, even a single body placement would likely suffice for locomotion activities. Detecting sedentary tasks had low accuracies, although the five-sensor combination provided a 7%-20% increase in balanced accuracy compared with several single placements. Additional analyses demonstrated that the misclassification rate was higher for sedentary activities than for nonsedentary activities. This may be caused by an imbalance in the data collected; sedentary tasks comprise only 4 out of 32 activities and result in only 6% of the total epochs. Another potential reason might be the similarity of some of the nonsedentary and sedentary activities. Confusion matrices of individual activity recognition models show that strength exercise and some stretching and some yoga, which were performed in a sitting position for a significant amount of time, contained most of the error (approximately 25%-76% for the hip and 40%-50% for the wrist). These activities are not traditionally considered to be sedentary behavior but are often performed in a sitting position (confusion matrices presented in Multimedia Appendices 2 and 3).

Historically, the hip position has been the most common and well-validated accelerometer placement. Some studies have investigated the performance of classifiers using data from other sensor placements, such as the ankle and wrist [22,25,49]. However, few studies have systematically examined the accuracy differences between individuals and combinations of different body placements [50,51]. The results published by Arif and Kattan [50] demonstrated in a cohort of 9 young adults that body placement differences between the wrist, chest, and ankle were relatively small in terms of overall accuracy when classifying 12 activities (best overall F-measure for wrist placement: 93.9%, for ankle placement: 92.2%, and for chest placement: 93.9% vs for combined placements: 98.2%). Similar findings have been reported by Gao et al [51], where the following 4 placement positions were compared: chest, underarm, waist, and thigh to identify 5 different activities performed by 8 older adults. They reported accuracies ranging from 81.9% to 92.8% for single-placement classifiers and 83.2%-96.4% for multisensor classifiers. These 2 studies were consistent with the finding that additional accelerometers improve performance in detecting the physical activity type. This study increases this initial knowledge with a much larger sample size of older adults who performed an ample number of activities with and without overlapping movement patterns. Although more generalizable, the large sample size likely introduced more variability in movement patterns, making it more challenging to find a single common classifier appropriate for all people. As such, the lower performance for activity recognition observed in this study might test the limits of the predictive capacity for machine learning models, such as random forest, when applied across a diverse population.

A MET RMSE of 0.88 was achieved across all activities. Previous studies using data from accelerometer devices worn on the hip and wrist have shown similar results for the prediction of METs, with RMSE values of 1.00-1.22 [45,52,53]. For a single placement, the hip and thigh positions provided the lowest RMSE values. Increasing the number of placements only slightly enhanced the RMSE (from 3% to 9%). Our results also show that adding 2 or more accelerometers provides a small enhancement in prediction. Previous studies with a smaller number of activities had similar performance in MET estimation—1.0 METs and 1.2 METs using data collected from wrist and hip placements [42,45]. Our slightly better performance might be because of a large range of activities that enhanced MET distribution.

We believe that our work constitutes one of the largest accelerometer-based validation studies in older adults. Data were collected at a high resolution, and there were a large number of activities included and 5 body placements. This resulted in a large number of pairwise (location and sensor) combinations. A limitation of this study is that data were collected in controlled laboratory settings, which is an appropriate initial step in a validation framework [54]. The next step is to collect data in free-living settings with more fluid transitions between tasks, which is more reflective of actual movement. Another limitation of the study was that not all activities were performed by all participants (Multimedia Appendix 1). However, the final number of participants with complete data for each activity was sufficient to assess the accuracy of individual body positions and their combinations. Another limitation of the study was that the performance ranking and conclusions were based on random forest models and might change when using other machine learning models. We used the random forest model because it was found to be the best performing in our previous study [41]. A subsequent analysis is required to validate whether the choice of machine learning model will affect the classification performance. Finally, our population included community-dwelling older volunteers to generalize to this population. Although this sample had common comorbidities such as diabetes, hypertension, and cancer history, we did not actively recruit people who had specific ambulatory deficits that would likely impact the results. Existing work in these specialized populations shows that knowledge from nonambulatory, impaired (eg, healthier) adults transfers with poor accuracy [55]. Thus, this study is limited to community-dwelling older adults without overt ambulatory deficits.

Conclusions

The results from this work suggest that additional accelerometer devices only slightly enhance activity recognition accuracy and MET estimation in older adults. However, no single or combination of accelerometer placement appeared to be significantly better than the others. Therefore, using a single accelerometer placement appears to provide sufficient

performance for labeling general activity categories and estimating energy expenditure. Researchers and practitioners should consider performance accuracy in the context of participant burden and the potential extra benefits gained in particular positions.

Acknowledgments

This research was primarily funded by the National Cancer Institute, grant SBIR HHSN261201500014C. This research was partially supported by the National Institute on Aging (grant R01AG042525) and the Claude D. Pepper Older Americans Independence Centers at the University of Florida (grant 1P30AG028740). PR was supported by the Career award, grant NSF-IIS 1750192, from the National Science Foundation, Division of Information and Intelligent Systems, and by grant NIH NIBIB R21EB027344-01. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Activity characteristics (four sedentary activities, six locomotion activities, and 22 lifestyle activities). RPE: rating of perceived exertion.
[DOCX File , 13 KB - mhealth_v9i5e23681_app1.docx]

Multimedia Appendix 2
Confusion matrix for individual activity recognition using data from hip placement.
[XLSX File (Microsoft Excel File), 15 KB - mhealth_v9i5e23681_app2.xlsx]

Multimedia Appendix 3
Confusion matrix for individual activity recognition using data from wrist placement.
[XLSX File (Microsoft Excel File), 15 KB - mhealth_v9i5e23681_app3.xlsx]

References

1. Gorman E, Hanson HM, Yang PH, Khan KM, Liu-Ambrose T, Ashe MC. Accelerometry analysis of physical activity and sedentary behavior in older adults: a systematic review and data analysis. Eur Rev Aging Phys Act 2014 Sep 17;11(1):35-49 [FREE Full text] [doi: 10.1007/s11556-013-0132-x] [Medline: 24765212]
2. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. Int J Behav Nutr Phys Act 2010 Aug 03;7(1):60 [FREE Full text] [doi: 10.1186/1479-5868-7-60] [Medline: 20682057]
3. Paudel ML, Taylor BC, Ancoli-Israel S, Stone KL, Tranah G, Redline S, et al. Rest/activity rhythms and cardiovascular disease in older men. Chronobiol Int 2011 Apr 31;28(3):258-266 [FREE Full text] [doi: 10.3109/07420528.2011.553016] [Medline: 21452921]
4. Berle JO, Hauge ER, Oedegaard KJ, Holsten F, Fasmer OB. Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. BMC Res Notes 2010 May 27;3(1):149 [FREE Full text] [doi: 10.1186/1756-0500-3-149] [Medline: 20507606]
5. Jones SH, Hare DJ, Evershed K. Actigraphic assessment of circadian activity and sleep patterns in bipolar disorder. Bipolar Disord 2005 Apr;7(2):176-186. [doi: 10.1111/j.1399-5618.2005.00187.x] [Medline: 15762859]
6. Berger AM, Wielgus K, Hertzog M, Fischer P, Farr L. Patterns of circadian activity rhythms and their relationships with fatigue and anxiety/depression in women treated with breast cancer adjuvant chemotherapy. Support Care Cancer 2010 Jan 19;18(1):105-114. [doi: 10.1007/s00520-009-0636-0] [Medline: 19381692]
7. Allenbach Y, Foucher A, Champtiaux N, Gilardin L, Hervier B, Benveniste O, et al. Wrist-worn accelerometer as innovative tool for longitudinal follow-up of idiopathic inflammatory myopathy patients: a pilot study. Neuromuscul Disord 2015 Oct;25:310. [doi: 10.1016/j.nmd.2015.06.442]
8. Ahanathapillai V, Amor JD, Tadeusiak M, James CJ. Wrist-worn accelerometer to detect postural transitions and walking patterns. In: Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing. 2013 Presented at: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013; September 25-28, 2013; Seville, Spain p. 1515-1518. [doi: 10.1007/978-3-319-00846-2_374]
9. Ekblom O, Nyberg G, Bak EE, Ekelund U, Marcus C. Validity and comparability of a wrist-worn accelerometer in children. J Phys Act Health 2012 Mar;9(3):389-393. [Medline: 22454440]

10. Lyden K, John D, Dall P, Granat MH. Differentiating sitting and lying using a thigh-worn accelerometer. Med Sci Sports Exerc 2016 Apr;48(4):742-747. [doi: 10.1249/MSS.0000000000000804] [Medline: 26516691]
11. Wullems JA, Verschueren SM, Degens H, Morse CI, Onambélé GL. Performance of thigh-mounted triaxial accelerometer algorithms in objective quantification of sedentary behaviour and physical activity in older adults. PLoS One 2017 Nov 20;12(11) [FREE Full text] [doi: 10.1371/journal.pone.0188215] [Medline: 29155839]
12. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Med Sci Sports Exerc 2013 May;45(5):964-975 [FREE Full text] [doi: 10.1249/MSS.0b013e31827f0d9c] [Medline: 23247702]
13. Kinder JR, Lee KA, Thompson H, Hicks K, Topp K, Madsen KA. Validation of a hip-worn accelerometer in measuring sleep time in children. J Pediatr Nurs 2012 Apr;27(2):127-133 [FREE Full text] [doi: 10.1016/j.pedn.2010.11.004] [Medline: 22341191]
14. Vähä-Ypyä H, Husu P, Suni J, Vasankari T, Sievänen H. Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer. Scand J Med Sci Sports 2018 Mar 13;28(3):1092-1102. [doi: 10.1111/sms.13017] [Medline: 29144567]
15. Chen H, Lin K, Hsieh Y, Wu C, Liing R, Chen C. A study of predictive validity, responsiveness, and minimal clinically important difference of arm accelerometer in real-world activity of patients with chronic stroke. Clin Rehabil 2018 Jan 05;32(1):75-83. [doi: 10.1177/0269215517712042] [Medline: 28580791]
16. Jakicic JM, Marcus M, Gallagher KI, Randall C, Thomas E, Goss FL, et al. Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. Med Sci Sports Exerc 2004 May;36(5):897-904. [doi: 10.1249/01.mss.0000126805.32659.43] [Medline: 15126727]
17. Anderson J, Green A, Hall H, Yoward S. Criterion validity of an ankle or waist mounted Actigraph GT3X accelerometer in measurement of body position and step count. Physiotherapy 2016 Nov;102:79-80. [doi: 10.1016/j.physio.2016.10.077]
18. Crouter SE, Oody JF, Bassett DR. Estimating physical activity in youth using an ankle accelerometer. J Sports Sci 2018 Oct 08;36(19):2265-2271 [FREE Full text] [doi: 10.1080/02640414.2018.1449091] [Medline: 29517959]
19. Lemmens RJ, Janssen-Potten YJ, Timmermans AA, Smeets RJ, Seelen HA. Recognizing complex upper extremity activities using body worn sensors. PLoS One 2015;10(3) [FREE Full text] [doi: 10.1371/journal.pone.0118642] [Medline: 25734641]
20. Pirttikangas S, Fujinami K, Nakajima T. Feature selection and activity recognition from wearable sensors. Ubiquitous Computing Systems. Berlin: Springer; 2006. URL: https://link.springer.com/content/pdf/10.1007%2F11890348.pdf [accessed 2021-03-29]
21. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical Human Activity Recognition Using Wearable Sensors. Sensors (Basel) 2015 Dec 11;15(12):31314-31338 [FREE Full text] [doi: 10.3390/s151229858] [Medline: 26690450]
22. Rhudy MB, Mahoney JM. A comprehensive comparison of simple step counting techniques using wrist- and ankle-mounted accelerometer and gyroscope signals. J Med Eng Technol 2018 Apr 30;42(3):236-243. [doi: 10.1080/03091902.2018.1470692] [Medline: 29846134]
23. Sandroff BM, Motl RW, Pilutti LA, Learmonth YC, Ensari I, Dlugonski D, et al. Accuracy of StepWatch™ and ActiGraph accelerometers for measuring steps taken among persons with multiple sclerosis. PLoS One 2014 Apr 8;9(4) [FREE Full text] [doi: 10.1371/journal.pone.0093511] [Medline: 24714028]
24. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: step counts at a wide range of gait velocities. Med Eng Phys 2014 Jun;36(6):659-669 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.006] [Medline: 24656871]
25. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. Med Sci Sports Exerc 2013 Nov;45(11):2193-2203 [FREE Full text] [doi: 10.1249/MSS.0b013e31829736d6] [Medline: 23604069]
26. Full KM, Kerr J, Grandner MA, Malhotra A, Moran K, Godoble S, et al. Validation of a physical activity accelerometer device worn on the hip and wrist against polysomnography. Sleep Health 2018 Apr;4(2):209-216 [FREE Full text] [doi: 10.1016/j.sleh.2017.12.007] [Medline: 29555136]
27. Beaver L. The Smartwatch Report: forecasts, adoption trends, and why the market isn't living up to the hype. businessinsider.com. 2016. URL: http://www.businessinsider.com/smartwatch-and-wearables-research-forecasts-trends-market-use-cases-2016-9 [accessed 2018-06-11] [WebCite Cache ID 705zlC4O2]
28. Smith A. Record shares of Americans now own smartphones, have home broadband. Pew Research Center. 2017. URL: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2018-06-11]
29. Kerr J, Marinac CR, Ellis K, Godbole S, Hipp A, Glanz K, et al. Comparison of accelerometry methods for estimating physical activity. Med Sci Sports Exerc 2017 Mar;49(3):617-624 [FREE Full text] [doi: 10.1249/MSS.0000000000001124] [Medline: 27755355]
30. Migueles JH, Cadenas-Sanchez C, Ekelund U, Nyström CD, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]

31. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):68-76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
32. Gorman M. Development and the rights of older people. In: The Ageing and Development Report. Oxfordshire, United Kingdom: Taylor & Francis; 1999:1-19.
33. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 1995 Mar 02;332(9):556-562. [doi: 10.1056/nejm199503023320902]
34. Corbett DB, Wanigatunga AA, Valiani V, Handberg EM, Buford TW, Brumback B, et al. Metabolic costs of daily activity in older adults (Chores XL) study: design and methods. Contemp Clin Trials Commun 2017 Jun;6:1-8 [FREE Full text] [doi: 10.1016/j.conctc.2017.02.003] [Medline: 28649668]
35. Time spent in detailed primary activities 1 and percent of the civilian population engaging in each detailed primary activity category, averages per day by sex, 2010 annual averages. American Time Use Survey, Bureau of Labor Statistics. URL: https://www.bls.gov/tus/tables/a1_2010.pdf [accessed 2021-03-25]
36. ActiGraph wGT3X-BT. Actigraphcorp. URL: https://actigraphcorp.com/actigraph-wgt3x-bt/ [accessed 2021-03-29]
37. Buchan DS, McSeveney F, McLellan G. A comparison of physical activity from Actigraph GT3X+ accelerometers worn on the dominant and non-dominant wrist. Clin Physiol Funct Imaging 2019 Jan 30;39(1):51-56. [doi: 10.1111/cpf.12538] [Medline: 30058765]
38. Dieu O, Mikulovic J, Fardy PS, Bui-Xuan G, Béghin L, Vanhelst J. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. Clin Physiol Funct Imaging 2017 Sep 07;37(5):525-529. [doi: 10.1111/cpf.12337] [Medline: 26749436]
39. McLaughlin JE, King GA, Howley ET, Bassett DR, Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. Int J Sports Med 2001 May;22(4):280-284. [doi: 10.1055/s-2001-13816] [Medline: 11414671]
40. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990 Aug;13(8):555-565 [FREE Full text] [doi: 10.1002/clc.4960130809] [Medline: 2204507]
41. Davoudi A, Wanigatunga AA, Kheirkhahan M, Corbett DB, Mendoza T, Battula M, et al. Accuracy of Samsung Gear S smartwatch for activity recognition: validation study. JMIR Mhealth Uhealth 2019 Feb 06;7(2):e11270 [FREE Full text] [doi: 10.2196/11270] [Medline: 30724739]
42. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. J Appl Physiol (1985) 2015 Aug 15;119(4):396-403 [FREE Full text] [doi: 10.1152/japplphysiol.00026.2015] [Medline: 26112238]
43. Mannini A, Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors (Basel) 2010 Feb 01;10(2):1154-1175 [FREE Full text] [doi: 10.3390/s100201154] [Medline: 22205862]
44. Breiman L. Random forests. In: Machine Learning. Switzerland: Springer Nature; 2001:5-32.
45. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov 23;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
46. Ho TK. Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition. 1995 Presented at: 3rd International Conference on Document Analysis and Recognition; August 14-16, 1995; Montreal, QC, Canada p. 278-282. [doi: 10.1109/icdar.1995.598994]
47. Velez DR, White BC, Motsinger AA, Bush WS, Ritchie MD, Williams SM, et al. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol 2007 May;31(4):306-315. [doi: 10.1002/gepi.20211] [Medline: 17323372]
48. Kuhn M. The caret package. R Foundation for Statistical Computing, Vienna, Austria. URL: https://cran.r-project.org/web/packages/caret/index.html [accessed 2021-03-29]
49. Thomas L, Khan U, Park Y, Ionita R, Jubran A, Ekekwe I, et al. Does wrist plus ankle actigraphy accurately quantify sleep architecture in patients weaning from prolonged mechanical ventilation? In: B56. Neurocognitive and Neuromuscular Issues in the ICU. 2020 Presented at: American Thoracic Society 2013 International Conference; May 17-22, 2013; Philadelphia Pennsylvania p. 3125 URL: https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2013.187.1_MeetingAbstracts.A3125
50. Arif M, Kattan A. Physical activities monitoring using wearable acceleration sensors attached to the body. PLoS One 2015 Jul 23;10(7) [FREE Full text] [doi: 10.1371/journal.pone.0130851] [Medline: 26203909]
51. Gao L, Bourke AK, Nelson J. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. Med Eng Phys 2014 Jun;36(6):779-785 [FREE Full text] [doi: 10.1016/j.medengphy.2014.02.012] [Medline: 24636448]
52. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol (1985) 2009 Oct;107(4):1300-1307 [FREE Full text] [doi: 10.1152/japplphysiol.00465.2009] [Medline: 19644028]

53. Montoye AH, Mudd LM, Biswas S, Pfeiffer KA. Energy expenditure prediction using raw accelerometer data in simulated free living. Med Sci Sports Exerc 2015 Aug;47(8):1735-1746. [doi: 10.1249/MSS.0000000000000597] [Medline: 25494392]
54. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev 2019 Oct;47(4):206-214. [doi: 10.1249/JES.0000000000000206] [Medline: 31524786]
55. Lonini L, Gupta A, Deems-Dluhy S, Hoppe-Ludwig S, Kording K, Jayaraman A. Activity recognition in individuals walking with assistive devices: the benefits of device-specific models. JMIR Rehabil Assist Technol 2017 Aug 10;4(2):e8 [FREE Full text] [doi: 10.2196/rehab.7317] [Medline: 28798008]

Abbreviations

CO_2: carbon dioxide
MET: metabolic equivalent
O_2: oxygen
RMSE: root mean square error
VO_2: oxygen consumption

Edited by L Buis; submitted 19.08.20; peer-reviewed by A Tarekegn, E Park; comments to author 29.09.20; revised version received 28.10.20; accepted 15.03.21; published 03.05.21.

Please cite as:
Davoudi A, Mardini MT, Nelson D, Albinali F, Ranka S, Rashidi P, Manini TM
The Effect of Sensor Placement and Number on Physical Activity Recognition and Energy Expenditure Estimation in Older Adults: Validation Study
JMIR Mhealth Uhealth 2021;9(5):e23681
URL: https://mhealth.jmir.org/2021/5/e23681
doi:10.2196/23681
PMID:33938809

Original Paper

Pain Assessment Tool With Electrodermal Activity for Postoperative Patients: Method Validation Study

Seyed Amir Hossein Aqajari[1], BSc; Rui Cao[1], BSc; Emad Kasaeyan Naeini[2], MSc; Michael-David Calderon[3], BS; Kai Zheng[4], PhD; Nikil Dutt[2], PhD; Pasi Liljeberg[5], PhD; Sanna Salanterä[6,7], PhD; Ariana M Nelson[3,8], MD; Amir M Rahmani[1,2,9,10], PhD

[1]Department of Electrical Engineering and Computer Science, University of California, Irvine, Irvine, CA, United States
[2]Department of Computer Science, University of California, Irvine, Irvine, CA, United States
[3]Department of Anesthesiology and Perioperative Care, University of California, Irvine, Irvine, CA, United States
[4]Department of Informatics, University of California, Irvine, Irvine, CA, United States
[5]Department of Future Technology, University of Turku, Turku, Finland
[6]Department of Nursing Science, University of Turku, Turku, Finland
[7]Turku University Hospital, University of Turku, Turku, Finland
[8]School of Medicine, University of California, Irvine, Irvine, CA, United States
[9]School of Nursing, University of California, Irvine, Irvine, CA, United States
[10]Institute for Future Health, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Seyed Amir Hossein Aqajari, BSc
Department of Electrical Engineering and Computer Science
University of California, Irvine
2200 Engineering Hall
Irvine, CA
United States
Phone: 1 9495379373
Email: saqajari@uci.edu

Abstract

Background: Accurate, objective pain assessment is required in the health care domain and clinical settings for appropriate pain management. Automated, objective pain detection from physiological data in patients provides valuable information to hospital staff and caregivers to better manage pain, particularly for patients who are unable to self-report. Galvanic skin response (GSR) is one of the physiologic signals that refers to the changes in sweat gland activity, which can identify features of emotional states and anxiety induced by varying pain levels. This study used different statistical features extracted from GSR data collected from postoperative patients to detect their pain intensity. To the best of our knowledge, this is the first work building pain models using postoperative adult patients instead of healthy subjects.

Objective: The goal of this study was to present an automatic pain assessment tool using GSR signals to predict different pain intensities in noncommunicative, postoperative patients.

Methods: The study was designed to collect biomedical data from postoperative patients reporting moderate to high pain levels. We recruited 25 participants aged 23-89 years. First, a transcutaneous electrical nerve stimulation (TENS) unit was employed to obtain patients' baseline data. In the second part, the Empatica E4 wristband was worn by patients while they were performing low-intensity activities. Patient self-report based on the numeric rating scale (NRS) was used to record pain intensities that were correlated with objectively measured data. The labels were down-sampled from 11 pain levels to 5 different pain intensities, including the baseline. We used 2 different machine learning algorithms to construct the models. The mean decrease impurity method was used to find the top important features for pain prediction and improve the accuracy. We compared our results with a previously published research study to estimate the true performance of our models.

Results: Four different binary classification models were constructed using each machine learning algorithm to classify the baseline and other pain intensities (Baseline [BL] vs Pain Level [PL] 1, BL vs PL2, BL vs PL3, and BL vs PL4). Our models achieved higher accuracy for the first 3 pain models than the BioVid paper approach despite the challenges in analyzing real patient data. For BL vs PL1, BL vs PL2, and BL vs PL4, the highest prediction accuracies were achieved when using a random

forest classifier (86.0, 70.0, and 61.5, respectively). For BL vs PL3, we achieved an accuracy of 72.1 using a k-nearest-neighbor classifier.

Conclusions: We are the first to propose and validate a pain assessment tool to predict different pain levels in real postoperative adult patients using GSR signals. We also exploited feature selection algorithms to find the top important features related to different pain intensities.

International Registered Report Identifier (IRRID): RR2-10.2196/17783

(JMIR Mhealth Uhealth 2021;9(5):e25258) doi:10.2196/25258

KEYWORDS

pain assessment; recognition; health monitoring; wearable electronics; machine learning; electrodermal activity; post-op patients

Introduction

Pain assessment is a key factor in successful pain management [1]. Inaccurate postoperative pain assessment may cause illnesses [2] and even long-term chronic issues [3]. Pain assessment tools for clinical use are in great demand. If communication ability is limited or even lost due to surgery or illness complications, it is difficult for a doctor to determine the patient's feelings. A proper pain assessment tool can offer an approximate pain level of that patient for further treatment. For now, the wide range of pain assessment methods still cannot determine the precise pain prevalence and levels for adults in hospitals [4]. That may cause incorrect treatment and lead to various problems and risks for patients. Painkillers have many side effects, and overtreatment of pain can trigger respiratory depression in the short term or substance use disorder in the long term [5]. However, undertreatment of pain may result in chronic pain, more health care costs, and physiological and psychological suffering [3,6]. All these issues mentioned are prevalent among noncommunicative patients [7]. A valid pain assessment tool would be truly transformative to health care delivery as clinicians could deliver pain treatments and assess response in real time. This would decrease unwanted side effects and recovery time from illness or a procedural intervention.

With the rapid development of Internet-of-Things (IoT) devices, including wearable devices, automated and continuous objective pain intensity assessment is possible [8]. The accuracy of these wearable devices has been evaluated in several studies. As an example, Mehrabadi et al [9] validated the accuracy of these devices in terms of sleep. Researchers try to identify nervous reaction to pain by monitoring the fluctuation in patients' physiological data, including electromyography, electrocardiography, photoplethysmography, and electrodermal activity (EDA) in real time [10,11]. Other research uses facial expression and head movement to accompany the physiological data [12,13]. These methods can quantify pain intensity, especially for poorly communicating patients [14]. However, all these methods to date use stimulated pain and were evaluated on healthy participants. Based on this observation, we developed the UCI iHurt Dataset (UCI_iHurtDB) [15]. The UCI_iHurtDB is the first multimodal dataset collected from postoperative adult patients suffering from real pain in hospitals. This dataset is planned to be released for research purposes in the near future.

Skin conductance or the EDA signal is considered a useful biomedical data point that corresponds to pain perception. Our skin produces sweat via over 3 million small tubular sweat glands. Sweat glands are distributed across the body but with the highest densities on the soles of the feet, palms and fingers, and forehead and cheeks. If a patient is exposed to a certain group of stimuli, they can be triggered to secrete moisture, termed emotional sweating. This results in a decrease in skin resistance, or in other words, an increase in skin conductance [16], which is also known as EDA or galvanic skin response (GSR). Other than pain, the EDA signal can extract a variety of valuable information from the human body. Rostami et al [17] highlighted an important insight that, by using the GSR signal, the biological impact of food on a person’s body can be captured.

Pain assessment research only using EDA data is limited. Eriksson et al [18] and Munsters et al [19] validated the relationship between EDA and pain for newborn infants, suitable for automated pain assessment due to their inability to communicate. By monitoring the EDA data during routine blood sampling or care intervention, they found EDA can differentiate between pain and no pain; however, more research is needed to achieve a clinical-grade level. Manivannan et al [20] verified whether the EDA could be used as a valid pain indicator for hypnotic analgesia with 10 participants. They used an iron disk to create mechanical pain in a laboratory setup. The experimental results show a clear relation between pain scores and EDA. None of these mentioned works used machine learning algorithms to create a classification model for pain assessment. Furthermore, their dataset includes healthy patients with various stimulus methods to cause pain. In another work, Susam et al [21] attempted to assess postoperative pain using EDA through a machine learning model. Their model could distinguish between clinical moderate-to-severe pain and no-pain conditions. However, their work only focused on children as a population.

To the best of our knowledge, for the first time in this paper, we present an automatic and versatile pain assessment tool to predict different pain levels in postoperative adults using only EDA signals as a physiological signal. In our pain assessment tool, we used 11 different time-domain features extracted from EDA signals for prediction. A feature selection algorithm is used to increase our tool's prediction accuracy and find the top-most important features related to pain intensity. To evaluate our results, we used different types of machine learning algorithms. Machine learning techniques and neural networks have been widely used in health monitoring domains. Zargari et al [22] used a combination of convolutional neural networks

and recurrent neural networks to accurately track the position of in-mouth nutrient sensors. Mehrabadi et al [23] used convolutional neural networks to detect COVID-19 in patients with acute respiratory distress syndrome. These techniques are not limited only to health care domains. Ashrafiamiri et al [24] used deep neural networks to secure autonomous driving. To accurately validate our pain assessment algorithm's performance, we compared our results with the accuracy achieved for the pain models presented in [13]. Werner et al [13] used the BioVid Heat Pain database in their work in which participants were subjected to painful heat stimuli under controlled settings. This comparison aims to show that despite all the challenges in real hospital settings, our models can achieve comparable results. This also shows the feasibility of using artificial intelligence–based objective pain assessment for real patients.

Methods

All methods of the study, including the data collection and pain assessment, were approved by the University of California Irvine Institutional Review Board (IRB, HS: 2017-3747). Potential candidates were screened for eligibility using the Acute Pain Service schedule and provided with a copy of the consent form to review for at least 24 hours before participation in research procedures.

Study Description, Participants, and Recruitment

This study is a biomedical data collection study with postoperative patients reporting varying degrees of pain symptoms under local IRB approval supervision. We recruited 25 participants (age: 23-89 years) from the University of California, Irvine Medical Center. We recruited similar numbers of men and women (13 men and 12 women). We removed 3 participants' data from the final dataset due to data recording accidents such as excessive motion artifacts induced by hand movements, and 2 participants' data were excluded since they were wearing the Empatica E4 watch on their IV arm, which resulted in unreliable EDA signal due to conditions such as skin rash and itching. The criteria for participant selection were as follows: (1) 18 years of age or older, (2) would receive a consult by the Acute Pain Service, (3) no barriers to communication, (4) able to provide written informed consent, and (5) have intact and healthy facial skin. Participants were excluded if they had any of the following: (1) any diagnosed condition affecting cognitive function like dementia or psychosis; (2) any diagnosed condition affecting the central nervous system, facial nerves, or muscles; (3) deformities on the hand or other parts of the body that prevent sensor placement; and (4) significant facial hair growth in the area for sensor attachment. Patients were selected if they satisfied the inclusion and exclusion criteria and determined to be enrolled in this study voluntarily. Verbal and written consent was acquired before initiation of the study.

Study Design

After the recruitment procedure, GSR data were collected for approximately 30 minutes continuously from the patients in their private room. We separated these 30 minutes of data monitoring into 2 phases. In the first phase, an artificial pain generator called a transcutaneous electrical nerve stimulation (TENS) [25] unit was used to let participants have an initial impression of multiple pain levels and let researchers obtain baseline biosignals from the person. The TENS unit stimulates different levels of acute pain by delivering small electrical impulses through electrodes attached to the participant's skin with adhesive pads. Participants were told to gradually increase the TENS unit's intensity to a tolerable level for them and then hold it for at least 10 seconds. After this, we decreased the intensity back to level 0. In the second phase, participants engaged in low-intensity activities such as walking, coughing, sitting up, or lifting legs that caused an expected degree of pain. To improve data reliability for the following analysis, the entire data monitoring process was repeated sequentially. The monitored person's self-report of pain was measured using the Numeric Rating Scale (NRS), a segmented numeric version of the Visual Analogue Scale (VAS). The VAS is a validated, subjective measure for acute and chronic pain. Pain scores are recorded by making a mark on a 10-cm line representing a sequence between "no pain" and "worst pain." NRS quantifies the pain intensity to 10 levels (0 is no pain, and integers 1 to 10 represent different pain levels, with 10 being the highest pain imaginable) [26,27].

Data Collection

We used the Empatica E4, the commercially available wristband, to monitor the EDA data. The wristband is simple to position, and participants can maneuver easily without the device impeding their movements in any way. The wristband's internal memory allows recording up to 36 hours of data and wireless data transmission. The E4 wristband is rechargeable, with a charging time of fewer than 2 hours. An EDA sensor is embedded in the E4 wristband. This sensor measures the fluctuating changes in certain electrical properties of the skin.

GSR Feature Extraction Pipeline Architecture

Figure 1 shows our pipeline architecture for preparing the data and extracting the set of features for classification. There are 3 different sections in this pipeline: (1) Data Preparation, (2) pyEDA [28], (3) Post Feature Extraction.

Figure 1. Galvanic skin response (GSR) feature extraction pipeline. EDA: electrodermal activity.

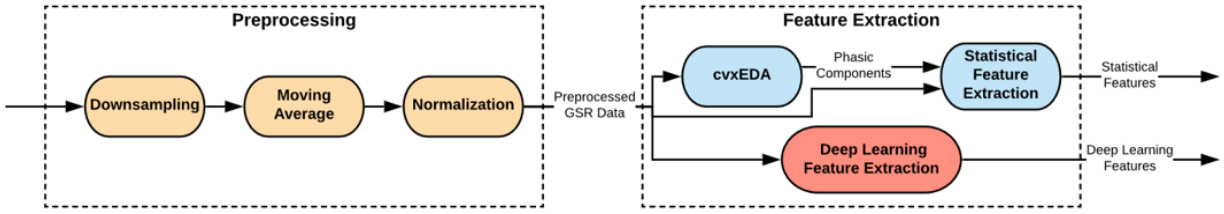
Data Preparation

The primary purpose of the Data Preparation in our pipeline is to synchronize the data with the labels. To prepare the data for feature extraction, we extracted the original signals' slices that match with their corresponding labels. With this aim, the slices of GSR data and their labels are collected in this part to be fed to the pyEDA for pre-processing and feature extraction.

pyEDA

The architecture of the pyEDA is shown in Figure 2. According to this figure, Preprocessing and Feature Extraction are the 2 main stages in this pipeline.

Figure 2. Pipeline architecture of the pyEDA. EDA: electrodermal activity; GSR: galvanic skin response.



In the preprocessing stage of the pyEDA pipeline, at first, the data are down-sampled; then, a moving average is used to smooth the data and reduce the artifacts such as body gestures and movements. In the end, the data are normalized to become suitable for classification models.

If the GSR data are collected at 128 Hz, it can safely be down-sampled to a 20 Hz sampling rate. This down-sampling has been done to conserve memory and processing time of the data. In this work, we did not down-sample the data since the original data are already sampled at 4 Hz, which is good in terms of time and memory usage.

In this work, several steps were taken to remove motion artifacts from the GSR signal. First, we used a moving average across a 1-second window to remove the motion artifacts and smooth the data. Second, a low-pass Butterworth filter on the phasic data was applied to remove the line noise. Lastly, preprocessed GSR signals corresponding to each different pain level were visualized to ensure the validity of the signals.

The pyEDA uses 2 different algorithms for feature extraction (Statistical Feature Extraction and Deep Learning Feature Extraction). The parameters of the Deep Learning Feature Extraction part of the pipeline are set and tuned for stress detection; therefore, in this work, we only used the features extracted by the Statistical Feature Extraction algorithm.

The number of peaks, the mean, and the max peak amplitude are the 3 different statistical features that are extracted in the pyEDA. The GSR signals consist of 2 main components: skin conductance level, also known as the tonic level of GSR, and skin conductance response, also called the phasic component of GSR. The GSR peaks or bursts are considered the variations in the phasic component of the signal. Therefore, the most important part in extracting the peaks of the GSR signal is to extract its phasic component. Based on Figure 2, the pyEDA tool uses the cvxEDA algorithm [29] to extract the phasic component. Then, the phasic component and the preprocessed GSR data are fed to the Statistical Feature Extraction module to extract the 3 mentioned features (number of peaks, mean GSR, and max peak amplitude).

Post Feature Extraction

We also extracted the features that were used in the work by Werner et al [13] for the GSR signals. The preprocessed GSR signals and the set of features (number of peaks, mean GSR, and max peak amplitude) were fed into the Post Feature Extraction module to extract these features.

The maximum value of the peaks, range, standard deviation, interquartile range, root mean square, mean value of local maxima, mean value of local minima, mean of the absolute values of the first differences, and mean of the absolute values of the second differences are the extra features that were extracted in this part. Table 1 shows all the extracted features with their descriptions.

The mean of the absolute values of the first differences (mavfd) is calculated as:

The mean of the absolute values of the second differences (mavsd) is calculated as:

Table 1. Extracted galvanic skin response (GSR) features with their descriptions.

Feature	Description
Number of peaks	The number of peaks
Mean	The mean value of the signal
Max	The maximum value of the peaks
Range	The difference between the maximum and the minimum value of the signal
STD	Standard deviation of the signal
IQR	The difference between upper and lower quartiles of the signal
RMS	Root mean square of the signal
Mean minima	The mean value of local minima of the signal
Mean maxima	The mean value of local maxima of the signal
mavfd	The mean of the absolute values of the first differences
mavsd	The mean of the absolute values of the second differences

Classification

Feature Selection

One of the key components in machine learning is to select the set of features that has the highest importance in classification. Performing feature selection on the data reduces overfitting, reduces training time, and improves accuracy. By removing the set of features that are not informative for our classification and only add complexity to our model, there is less opportunity to make decisions based on noise, making the model less over-fitted. Fewer data means less training time. In the end, by having more informative data and fewer misleading data, the accuracy of the model increases.

Random forests [30] are among the most popular machine learning methods. They provide 2 methods for feature selection: mean decrease impurity and mean decrease accuracy. In this work, we used a mean decrease impurity method for feature selection.

Mean decrease impurity is also sometimes called Gini importance. Random forest is an ensemble learning algorithm consisting of several decision trees. The decision tree is a tree-like model of decisions in which every node is a condition on one of the features. These nodes separate the data into 2 different sets so that in the optimal scenario, the data with the same labels end up in the same set. Impurity is the measure based on which the optimal condition is chosen on every node. Mean decrease impurity for each feature is defined as the total decrease in node impurity averaged over all ensemble trees. The features are ranked according to this measure.

Labeling the Features

Table 2 shows the distribution of labels in the UCI_iHurtDB. There exists 11 different pain levels in this dataset. It is noticeable that the distribution of different pain levels for the patients is imbalanced (4 occurrences of pain level 10, but 83 occurrences of pain level 4, as an example). This is understandable due to the subjective nature and the different sources of pain among our patients.

Table 2. Distribution of labels in the UCI_iHurtDB before down-sampling.

Pain level	Frequency, n
PL0	37
PL1	52
PL2	37
PL3	61
PL4	83
PL5	44
PL6	32
PL7	16
PL8	46
PL9	26
PL10	4

In the work by Werner et al [13], there were 5 different pain levels, including the baseline level. To properly compare our pain assessment algorithm with their work, we down-sampled our 11 classes to 5 classes. The key factor in this down-sampling

is to ensure that the distribution of the labels is as balanced as possible. As a result, we considered pain levels 1-3 as new pain level 1 (PL1), pain level 4 as new pain level 2 (PL2), pain levels 5-7 as new pain level 3 (PL3), and pain levels 8-10 as new pain level 4 (PL4). Based on Table 2, there are only 37 data points for the baseline. To increase the number of samples for the baseline to make our labels more balanced, we up-sampled PL0 based on the reported PL0 data by the patients. We ensured these new baseline data were close enough to the reported pain level 0 labels (less than 10 seconds difference) and had no overlap with other labels. These assumptions were made to make sure (1) we were not reproducing any data and (2) the patients had the same pain level 0 for these new timestamps. By doing this procedure for all the participants, our number of samples for pain level 0 increased from 37 to 86.

Table 3 shows the distribution of the down-sampled labels and the new baseline. The distribution of the new labels is appropriately balanced. Still, for PL1, the number of samples is slightly higher than the rest of the classes. This is because we down-sampled our pain levels to 4 different classes to make our settings comparable with the work by Werner et al [13].

Table 3. Distribution of labels in the UCI_iHurtDB after down-sampling.

Pain level	Frequency, n
PL0	86
PL1	150
PL2	83
PL3	92
PL4	76

Machine Learning Algorithms

We used machine learning–based algorithms to evaluate the performance of our pain assessment algorithm. Two different classification methods were used here: (1) k-nearest-neighbor with *k* between 1 and 20 and (2) random forest with a depth between 1 and 10. The k-nearest-neighbor method uses *k* number of nearest data points and predicts the result based on a majority vote [31]. The random forest classifier is an ensemble learning algorithm that fits several decision tree classifiers on various subsamples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [30]. We used the Scikit-learn library to create our classification models [32]. Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression, and clustering algorithms, including k-nearest-neighbor and random forest.

To accurately evaluate the performance of our classification models, we used a cross-validation method [33]. Cross-validation is one of the most popular algorithms used to truly estimate a machine learning model's accuracy on unseen data. It achieves this by training a model using different subsets of data and obtaining the average accuracy on the rest of the data as a test. In this work, we used leave-one-out cross-validation to evaluate our result. We considered all the data acquired from one of the patients as a test and created our pain model using the rest of the patients. We repeated this procedure for each patient as a test. Each time, we created our pain model from scratch without considering the current test patient data or any information from the previous pain models. The final accuracy of the model was obtained by averaging the accuracy of all constructed pain models.

Results

A total of 25 patients with acute pain were engaged by the Acute Pain Service and recruited for this study. Of these 25 participants, 5 were removed from our study due to problems in the data collection process due to rapid hand movements or unreliable EDA signal due to conditions such as skin rash and itching resulting from wearing the Empatica E4 watch on their IV arm. The average age of patients in this study was 54.45 (SD 17.44, range 23-89) years; 55% (11/20) of the patients were male, and 45% (9/20) of patients were female. All patients were taking the standard-of-care postoperative pain medications at the time of the study. Enrolled participants agreed to perform the research protocol for a median of 4 (IQR 3-6) days after surgical intervention. The nature of the procedures for each participant included the following domains: 45% (9/20) general surgery (diagnostic laparoscopy, exploratory laparotomy, and vascular), 15% (3/20) trauma (thoracic pain and rib plating), and 10% (2/20) urology (cystectomy and bladder augmentation). Also, 40% (8/20) of enrolled participants received standard-of-care epidural analgesia provided by the acute pain service team at the time of research participation. The remaining participants were receiving oral and intravenous analgesics for pain control (Table 4).

Table 4. Summary of patients' demographic characteristics for this study including exclusions.

Variable	Value	Range
Patients excluded for hand movement noise, n (%)	3 (12)	N/A[a]
Patients excluded for IV arm effect, n (%)	2 (8)	N/A
Age (years), mean (SD)	54.45 (17.44)	23-89
Gender, male, n (%)	11 (55)	N/A
Weight (kg), mean (SD)	75.24 (14.60)	52.2-102
Height (cm), mean (SD)	170.07 (9.00)	154.9-185.9
BMI (kg/m^2), mean (SD)	26.21 (5.75)	15.1-38.6
Nature of the procedure, n (%)		
General surgery	9 (45)	N/A
Orthopedics	6 (30)	N/A
Trauma	3 (15)	N/A
Urology	2 (10)	N/A

[a]N/A: not applicable.

To show that our pain assessment algorithm can achieve comparable results to the work by Werner et al [13], we used identical settings as their work. Werner et al [13] used 5 different pain levels, including the baseline. They also considered 5.5-second windows for the GSR data. Therefore, in the Data Preparation part of our pipeline, we considered 5.5-second windows for the slices of the GSR data for feature extraction (2.75 seconds before and after each timestamp). Furthermore, as discussed in the Methods section, we down-sampled the pain levels from 11 classes to 5 classes to make them similar with their labels.

At first, we used the set of features that was used in the work by Werner et al [13] for classification without any feature selection. The maximum value of the peaks, range, standard deviation, interquartile range, root mean square, mean value of local maxima, mean value of local minima, mean of the absolute values of the first differences, and mean of the absolute values of the second differences are the features that were used here.

We used leave-one-person-out cross-validation using k-nearest-neighbor and random forest algorithms. We reported the accuracy based on 4 different pain models (BL vs PL1, BL vs PL2, BL vs PL3, and BL vs PL4). Table 5 shows the comparison of the validation accuracy achieved by our classifiers with that by the pain models of Werner et al [13].

Table 5. The validation accuracies in comparison with Werner et al [13] using the same set of features.

Binary classification	RF[a]	KNN[b]	Werner et al [13]
BL vs PL1	84.0	74.4	55.4
BL vs PL2	66.3	67.5	60.2
BL vs PL3	57.2	65.0	65.9
BL vs PL4	55.2	53.0	73.8

[a]RF: random forest.

[b]KNN: k-nearest-neighbor.

According to these data, for the first 2 pain models (BL vs PL1 and BL vs PL2), we achieved a higher accuracy using both of our classifiers in comparison with Werner et al [13]. For the third pain model, our accuracy is also close to their models, with less than 1% difference using the k-nearest-neighbor classifier. For the fourth model, the accuracy of our models was noticeably lower than their models. As the next step, we added 2 more features (the number of peaks and the mean of the GSR data) to our set of features and then selected the most important ones using the mean decrease impurity method to improve the accuracy.

To obtain the best set of classification features, we ran leave-one-person-out cross-validation on different pain models using random forest classifiers. We computed the Gini importance of the features on the training data and selected the top *k* number of features for training the model and classification (2-11 were considered to be possible values for *k*). Since we had a different number of folds, we could have different sets of features for each fold. We considered the set of features that was used in most of the folds as the final set of features for the current pain model. Table 6 shows the selected features for each pain model. Descriptions for each of these features can be found in Table 1.

Table 6. Selected set of features for each pain model.

Pain models	Set of features
BL vs PL1	Mean, max, RMS[a], and mean maxima
BL vs PL2	Mean, max, RMS, and mean maxima
BL vs PL3	Max and RMS
BL vs PL4	Mean, max, IQR, RMS, and mean maxima

[a]RMS: root mean square.

According to this table, the maximum value and root mean square of the signal are 2 features that were selected in all the pain models. The mean value of the local maxima and the mean value of the signal were also selected for all the pain models except the third one. The difference between upper and lower quartiles of the signal (IQR) is a feature that was selected as an important feature for classification only for the BL vs PL4 pain model.

After the set of features for each pain model were obtained, we ran one-person-leave-out cross-validation for k-nearest neighbor and random forest algorithms using these sets of new selected features to achieve the final results.

Table 7 shows the validation accuracy comparison of our models with those by Werner et al [13] using feature selection for 5.5-second windows. As shown in this table, by using feature selection, our classifiers' accuracy improved compared to those shown in Table 5. For all the pain models except the fourth one, we were able to achieve higher accuracy than the pain models by Werner et al [13]. For BL vs PL4, our accuracy was still about 10% less than their work. In the Discussion section, we explain the potential reasons for this difference in our model.

Table 7. Validation accuracies in comparison with the work by Werner et al [13] using feature selection.

Binary classification	RF[a]	KNN[b]	Werner et al [13]
BL vs PL1	86.0	76.8	55.4
BL vs PL2	70.0	69.1	60.2
BL vs PL3	69.8	72.1	65.9
BL vs PL4	61.5	60.0	73.8

[a]RF: random forest.

[b]KNN: k-nearest-neighbor.

Discussion

Strengths

We are the first to develop an automatic and versatile pain assessment tool using GSR signals to accurately predict different pain intensities in postoperative adult patients. According to our results, using identical settings and even the exact GSR features used in the work by Werner et al [13], we can achieve higher accuracy in 2 of the 4 different pain models (BL vs PL1 and BL vs PL2). Also, for the third pain model (BL vs PL3), using the k-nearest-neighbor classifier, we can achieve the same level of accuracy with less than 1% difference. Machine learning algorithms and feature selection are not a part of the pain assessment algorithm settings. As a result, to show the true strength of our pain assessment algorithm, first, we added 2 more features to our set of features (the number of peaks and the mean value of the signal). Then, we used the mean decrease impurity method to select the most important features for classification. According to Table 7, we can achieve higher accuracy for the first 3 pain models using this procedure. Based on our results, for the first pain model (BL vs PL1), our accuracy is considerably higher than the accuracy achieved by Werner et al [13] (with and without feature selection). By feature selection, we are using a much lower number of features in our pain models. This reduces the complexity of our pain models.

Furthermore, we present the most important features for each classification model. According to Table 5, the maximum value and root mean square of the signal appear to be the 2 most important features for pain classifications for postoperative adult patients. The mean value and mean value of the local maxima of the signal are considered to be the next 2 important features for classification.

Our results show that, in addition to healthy participants from previous studies, we can detect different intensities of pain using only GSR data in real-life patients in the hospital. GSR data can easily be collected using affordable wearable devices such as the Empatica E4 used in our study. Therefore, our pain assessment algorithm is really beneficial in providing valuable information to hospital staff and caregivers to better manage pain, especially for those patients who cannot communicate.

Limitations

As our work's limitations, we used data collected from postoperative adult participants in our pain assessment algorithm. Data collection in real life may have led to more motion artifact noise in our physiological signals than data collection in a lab setting (especially for GSR data collected from Empatica E4 connected to their wrist). The motion artifacts were stronger for higher pain levels since the patient was in more unbearable pain.

Another limitation in our work can be the lack of balanced pain levels for all the patients. Since our data were collected from real postoperative adult participants, it is possible that the patients in the experiment did not experience and report all different pain levels. This limitation could be more noticeable at higher pain levels.

Based on our results, we were not able to achieve the accuracy of that by Werner et al [13] for BL vs PL4 (with a 10% difference). The limitations mentioned in previous paragraphs can explain the potential reasons behind this difference.

Furthermore, we could not find a significant difference between different pain levels in our study. We believe this is because the variations in GSR signal response to different pain levels are more alike to be distinguished easily. It is worth mentioning that most state-of-the-art pain assessments mainly focus on comparing baseline with other pain levels (eg, Werner et al [13], if the patient is in pain or not).

According to our results, the accuracy of our pain models generally decreases with increasing pain levels. This is the opposite of the models from Werner et al [13]. We believe this is due to our work's limitations mentioned in the earlier paragraph (motion artifact in signals and lack of balanced pain levels for all the patients). Both of these mentioned limitations could be more noticeable at higher pain levels. Therefore, it is understandable that in this work, where the data were collected from postoperative patients in real life, our models' accuracy decreased with increasing pain levels.

Conclusions

In conclusion, according to our results, we can evaluate the performance of our pain assessment algorithm. This evaluation shows that it is feasible to predict different pain levels in real postoperative adult participants using only the EDA data. Furthermore, we showed that the mean value, maximum value, root mean square, and mean value of the local maxima of the signal are the most important features for pain classification of real patients in pain. Multimodel pain assessment methods can be implemented as future work to increase our pain assessment models' performance.

Acknowledgments

This research was supported by the Academy of Finland project, Personalized Pain Assessment System based on IoT 313488.

Conflicts of Interest

None declared.

References

1. Serpell M. Handbook of pain management. New York, NY: Springer Healthcare; 2008.
2. Tompkins DA, Hobelmann JG, Compton P. Providing chronic pain management in the "Fifth Vital Sign" Era: Historical and treatment perspectives on a modern-day medical dilemma. Drug Alcohol Depend 2017 Apr 01;173 Suppl 1:S11-S21 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.12.002] [Medline: 28363315]
3. McGreevy K, Bottros M, Raja S. Preventing Chronic Pain following Acute Pain: Risk Factors, Preventive Strategies, and their Efficacy. Eur J Pain Suppl 2011 Nov 11;5(2):365-372 [FREE Full text] [doi: 10.1016/j.eujps.2011.08.013] [Medline: 22102847]
4. Gregory J, McGowan L. An examination of the prevalence of acute pain for hospitalised adult patients: a systematic review. J Clin Nurs 2016 Mar;25(5-6):583-598. [doi: 10.1111/jocn.13094] [Medline: 26778249]
5. McQuay H, Moore A, Justins D. Treating acute pain in hospital. BMJ 1997 May 24;314(7093):1531-1535 [FREE Full text] [doi: 10.1136/bmj.314.7093.1531] [Medline: 9183203]
6. Stites M. Observational pain scales in critically ill adults. Crit Care Nurse 2013 Jun;33(3):68-78. [doi: 10.4037/ccn2013804] [Medline: 23727853]
7. Barr J, Fraser GL, Puntillo K, Ely EW, Gélinas C, Dasta JF, American College of Critical Care Medicine. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013 Jan;41(1):263-306. [doi: 10.1097/CCM.0b013e3182783b72] [Medline: 23269131]
8. Naeini EK, Shahhosseini S, Subramanian A, Yin T, Rahmani A, Dutt N. An Edge-Assisted and Smart System for Real-Time Pain Monitoring. 2019 Presented at: 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); September 25-27, 2019; Arlington, VA. [doi: 10.1109/chase48038.2019.00023]
9. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep Tracking of a Commercially Available Smart Ring and Smartwatch Against Medical-Grade Actigraphy in Everyday Settings: Instrument Validation Study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465 [FREE Full text] [doi: 10.2196/20465] [Medline: 33038869]
10. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue H. Towards Pain Monitoring: Facial Expression, Head Pose, a new Database, an Automatic System and Remaining Challenges. 2014 Presented at: British Machine Vision Conference 2013; September 9-13, 2013; Bristol, England. [doi: 10.5244/c.27.119]
11. Walter S, Gruss S, Ehleiter H, Tan J, Traue H, Crawcour S, et al. The biovid heat pain database: Data for the advancement and systematic validation of an automated pain recognition. 2013 Presented at: 2013 IEEE International Conference on Cybernetics (CYBCO); June 13-15, 2013; Lausanne, Switzerland. [doi: 10.1109/cybconf.2013.6617456]

12. Hammal Z, Cohn JF. Automatic detection of pain intensity. 2012 Presented at: 14th ACM international conference on Multimodal interaction; October 2012; Santa Monica, CA. [doi: 10.1145/2388676.2388688]
13. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue HC. Automatic pain recognition from video and biomedical signals. 2014 Presented at: 22nd International Conference on Pattern Recognition; August 24-28, 2014; Stockholm, Sweden. [doi: 10.1109/icpr.2014.784]
14. Herr K, Coyne PJ, McCaffery M, Manworren R, Merkel S. Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations. Pain Manag Nurs 2011 Dec;12(4):230-250. [doi: 10.1016/j.pmn.2011.10.002] [Medline: 22117755]
15. Kasaeyan Naeini E, Jiang M, Syrjälä E, Calderon M, Mieronkoski R, Zheng K, et al. Prospective Study Evaluating a Pain Assessment Tool in a Postoperative Environment: Protocol for Algorithm Testing and Enhancement. JMIR Res Protoc 2020 Jul 01;9(7):e17783 [FREE Full text] [doi: 10.2196/17783] [Medline: 32609091]
16. Galvanic Skin Response - The Complete Pocket Guide. IMotions. 2017. URL: https://imotions.com/guides/eda-gsr/ [accessed 2021-04-11]
17. Rostami A, Pandey V, Nag N, Wang V, Jain R. Personal Food Model. USA: Association for Computing Machinery; 2020 Presented at: 28th ACM International Conference on Multimedia; October 2020; Seattle, WA. [doi: 10.1145/3394171.3414691]
18. Eriksson M, Storm H, Fremming A, Schollin J. Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. Acta Paediatr 2008 Jan;97(1):27-30. [doi: 10.1111/j.1651-2227.2007.00586.x] [Medline: 18052991]
19. Munsters J, Wallström L, Agren J, Norsted T, Sindelar R. Skin conductance measurements as pain assessment in newborn infants born at 22-27 weeks gestational age at different postnatal age. Early Hum Dev 2012 Jan;88(1):21-26 [FREE Full text] [doi: 10.1016/j.earlhumdev.2011.06.010] [Medline: 21764228]
20. Shankar K, Manivannan M, Fernandez P. GSR analysis of hypnotic analgesia. International Journal of Systemics, Cybernetics and Informatics 2008:35-38. [doi: 10.1016/b978-012370880-9.00212-7]
21. Susam B, Akcakaya M, Nezamfar H, Diaz D, Xu X, de Sa VR, et al. Automated Pain Assessment using Electrodermal Activity Data and Machine Learning. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:372-375 [FREE Full text] [doi: 10.1109/EMBC.2018.8512389] [Medline: 30440413]
22. Zargari AHA, Dautta M, Ashrafiamiri M, Seo M, Tseng P, Kurdahi F. NEWERTRACK: ML-Based Accurate Tracking of In-Mouth Nutrient Sensors Position Using Spectrum-Wide Information. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst 2020 Nov;39(11):3833-3841. [doi: 10.1109/tcad.2020.3013074]
23. Mehrabadi MA, Aqajari SAH, Azimi I, Downs CA, Dutt N, Rahmani AM. Detection of COVID-19 Using Heart Rate and Blood Pressure: Lessons Learned from Patients with ARDS. arXiv Prepr arXiv201110470 2020 2020:1-7 [FREE Full text]
24. Ashrafiamiri M, Pudukotai DS, Afandizadeh ZA, Seo M, Kurdahi F, Homayoun H. R2AD: Randomization and reconstructor-based adversarial defense for deep neural networks. 2020 Presented at: MLCAD 2020: 2020 ACM/IEEE Workshop on Machine Learning for CAD; November 2020; Virtual. [doi: 10.1145/3380446.3430628]
25. HealthmateForever Therapeutic Solutions. URL: https://www.healthmateforever.com/products/fda-cleared-otc-pain-relief-yk15ab-tens-units [accessed 2021-04-11]
26. Thong ISK, Jensen MP, Miró J, Tan G. The validity of pain intensity measures: what do the NRS, VAS, VRS, and FPS-R measure? Scand J Pain 2018 Jan 26;18(1):99-107. [doi: 10.1515/sjpain-2018-0012] [Medline: 29794282]
27. Johnson C. Measuring Pain. Visual Analog Scale Versus Numeric Pain Scale: What is the Difference? Journal of Chiropractic Medicine 2005 Dec;4(1):43-44. [doi: 10.1016/s0899-3467(07)60112-8]
28. Aqajari S, Naeini E, Mehrabadi M, Labbaf S, Rahmani A, Dutt N. GSR Analysis for Stress: Development and Validation of an Open Source Tool for Noisy Naturalistic GSR Data. arXiv:2005.01834 2020:1 [FREE Full text]
29. Greco A, Valenza G, Lanata A, Scilingo E, Citi L. cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing. IEEE Trans. Biomed. Eng 2016:1-1. [doi: 10.1109/tbme.2015.2474131]
30. Breiman L. Random forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
31. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. The American Statistician 1992 Aug;46(3):175. [doi: 10.2307/2685209]
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
33. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. 1995 Presented at: 14th International Joint Conference on Artificial Intelligence; August 1995; San Francisco, CA.

Abbreviations

EDA: electrodermal activity
GSR: galvanic skin response
IoT: Internet of Things
IRB: institutional review board

NRS: Numeric Rating Scale
TENS: transcutaneous electrical nerve stimulation
VAS: Visual Analogue Scale

Edited by L Buis; submitted 25.10.20; peer-reviewed by A Aminifar, I Shubina, N Taneja, D Firl; comments to author 23.11.20; revised version received 18.01.21; accepted 25.03.21; published 05.05.21.

Please cite as:
Aqajari SAH, Cao R, Kasaeyan Naeini E, Calderon MD, Zheng K, Dutt N, Liljeberg P, Salanterä S, Nelson AM, Rahmani AM
Pain Assessment Tool With Electrodermal Activity for Postoperative Patients: Method Validation Study
JMIR Mhealth Uhealth 2021;9(5):e25258
URL: https://mhealth.jmir.org/2021/5/e25258
doi:10.2196/25258
PMID:33949957

Original Paper

Acute Exacerbation of a Chronic Obstructive Pulmonary Disease Prediction System Using Wearable Device Data, Machine Learning, and Deep Learning: Development and Cohort Study

Chia-Tung Wu[1*], MSc; Guo-Hung Li[2*], MSc; Chun-Ta Huang[3], MD, PhD; Yu-Chieh Cheng[2], BSc; Chi-Hsien Chen[4], MD, PhD; Jung-Yien Chien[3], MD, PhD; Ping-Hung Kuo[3], MD; Lu-Cheng Kuo[3], MD; Feipei Lai[1], PhD

[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[2]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan

[3]Department of Internal Medicine, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan

[4]Department of Environmental and Occupational Medicine, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Jung-Yien Chien, MD, PhD
Department of Internal Medicine
National Taiwan University Hospital, College of Medicine, National Taiwan University
No 7 Chung-Shan S Road
Taipei, 100
Taiwan
Phone: 886 972651516
Email: jychien@ntu.edu.tw

Abstract

Background: The World Health Organization has projected that by 2030, chronic obstructive pulmonary disease (COPD) will be the third-leading cause of mortality and the seventh-leading cause of morbidity worldwide. Acute exacerbations of chronic obstructive pulmonary disease (AECOPD) are associated with an accelerated decline in lung function, diminished quality of life, and higher mortality. Accurate early detection of acute exacerbations will enable early management and reduce mortality.

Objective: The aim of this study was to develop a prediction system using lifestyle data, environmental factors, and patient symptoms for the early detection of AECOPD in the upcoming 7 days.

Methods: This prospective study was performed at National Taiwan University Hospital. Patients with COPD that did not have a pacemaker and were not pregnant were invited for enrollment. Data on lifestyle, temperature, humidity, and fine particulate matter were collected using wearable devices (Fitbit Versa), a home air quality–sensing device (EDIMAX Airbox), and a smartphone app. AECOPD episodes were evaluated via standardized questionnaires. With these input features, we evaluated the prediction performance of machine learning models, including random forest, decision trees, k-nearest neighbor, linear discriminant analysis, and adaptive boosting, and a deep neural network model.

Results: The continuous real-time monitoring of lifestyle and indoor environment factors was implemented by integrating home air quality–sensing devices, a smartphone app, and wearable devices. All data from 67 COPD patients were collected prospectively during a mean 4-month follow-up period, resulting in the detection of 25 AECOPD episodes. For 7-day AECOPD prediction, the proposed AECOPD predictive model achieved an accuracy of 92.1%, sensitivity of 94%, and specificity of 90.4%. Receiver operating characteristic curve analysis showed that the area under the curve of the model in predicting AECOPD was greater than 0.9. The most important variables in the model were daily steps walked, stairs climbed, and daily distance moved.

Conclusions: Using wearable devices, home air quality–sensing devices, a smartphone app, and supervised prediction algorithms, we achieved excellent power to predict whether a patient would experience AECOPD within the upcoming 7 days. The AECOPD prediction system provided an effective way to collect lifestyle and environmental data, and yielded reliable predictions of future AECOPD events. Compared with previous studies, we have comprehensively improved the performance of the AECOPD prediction model by adding objective lifestyle and environmental data. This model could yield more accurate prediction results for COPD patients than using only questionnaire data.

(JMIR Mhealth Uhealth 2021;9(5):e22591) doi:10.2196/22591

KEYWORDS

chronic obstructive pulmonary disease; clinical decision support systems; health risk assessment; wearable device

Introduction

With rapid progress of medicine, many treatments and medications have been developed, and relationships between lifestyle and disease have been elucidated. Precision medicine involves determining the best treatment plan for individual patients. Currently, research on precision medicine primarily involves developing related apps based on historical data from electronic medical records. When a patient is discharged from the hospital, lifestyle and environmental risks affect disease control. However, such factors are difficult to collect and use for analysis.

The World Health Organization includes chronic respiratory diseases among the four major human chronic diseases; in particular, lung disease accounts for an estimated 7.5 million deaths per year, or approximately 14% of annual deaths worldwide. These diseases are a major economic burden, and contribute to gender and social inequalities within and between countries. In descending frequency, the most frequent diseases include chronic obstructive pulmonary disease (COPD), lung cancer, tuberculosis, lung infections, asthma, and interstitial lung diseases [1]. COPD is a highly prevalent lung disease characterized by persistent airflow limitation due to a mixture of obstructive bronchiolitis and emphysema. The morbidity and mortality of COPD are high and continue to increase [2], such that COPD is projected to become the third-leading cause of death worldwide by 2030.

Acute exacerbation of COPD (AECOPD) decreases the patient's quality of life, accelerates decline in lung function, and is significantly associated with mortality [3]. COPD is a heterogeneous disorder with large variations in the risk of exacerbation across patients. In clinical practice, a history of two or more exacerbations and one severe exacerbation per year is used to guide therapeutic choices for exacerbation prevention [3]. However, this approach is clinically limited owing to significant heterogeneity in risk even among those who have frequent exacerbation episodes. Although these outcomes may be avoided with early detection and treatment, increasing evidence shows that environmental and lifestyle factors may affect the development of COPD.

Lifestyle modification is considered to be one of the most cost-effective strategies in the self-management and secondary prevention of COPD [4]. Nevertheless, there is limited evidence demonstrating the relationship between lifestyle factors and COPD development. Several studies have developed predictive models for AECOPD [5]; however, there is no prediction tool incorporating both lifestyle data and medical questionnaires. Moreover, some researchers have argued that remote monitoring is a promising alternative—or an adjunct—to traditional health care services in COPD management [6]. Nonetheless, some studies have shown that inefficient systems, poor patient compliance, and poor performance of prediction tools may decrease the effects of health care interventions [7-9].

Hence, the objectives of this study were to (1) develop a lifestyle observation platform based on wearable devices and a smartphone app to observe lifestyle and environmental factors for patients with COPD, and to (2) construct an AECOPD prediction tool for the early prediction of COPD exacerbations using lifestyle factors, indoor environmental factors, and medical questionnaires.

Methods

Data Collection

Eligible participants were adult patients with COPD (20 years of age or older) who were not implanted with a pacemaker and who were not pregnant. Participants were recruited from the pulmonologist clinics at National Taiwan University Hospital between March 2019 and February 2020. The study protocol was approved by the institutional review board of National Taiwan University Hospital (201710066RINB). During the study period, we enrolled 67 patients with a confirmed diagnosis of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [3], as defined by the ratio of <70% postbronchodilator forced expiratory volume in 1 second to the forced vital capacity.

Data collection was based on clinical questionnaires, environmental data, and physiological data. Data on patient symptoms were prospectively evaluated by the modified Medical Research Council (mMRC) dyspnea scale and the COPD assessment test (CAT) upon enrollment and every month during follow-up. The mMRC scale is used to assess functional impairment due to dyspnea attributable to respiratory disease, and the CAT is a patient-completed questionnaire that is used globally to assess the impact of COPD (cough, sputum, dyspnea, chest tightness) on health status. Both are widely used clinical tests for COPD, and some studies [4,10] have also used these clinical tools to evaluate the health condition of COPD patients. According to the GOLD guidelines, COPD exacerbations are defined as the acute worsening of respiratory symptoms, resulting in additional therapy [11].

Environmental data and physiological data were collected as time-series data by home air quality–sensing devices and wearable devices that were provided to all participants. Although most medical research [4,10,12] tends to use medical devices as data sources, patient compliance could be reduced by the inconvenience and difficulty in operating such devices. Physiological data included walking steps, climbing stairs, distances, consumption in calories, heart rate, and sleep status. Environmental data comprised temperature, humidity, and fine particulate matter (PM2.5) levels. All data were synchronized with the database every 15 minutes to ensure that any subtle changes were not neglected.

System Architecture

Figure 1 shows the architecture of the AECOPD prediction system. The system consists of four components: a wearable device (Fitbit Versa), home air quality–sensing device (EDIMAX Airbox), lifestyle observation platform, and the personal health advice app. Wearable devices automatically collected lifestyle data (physical activities, heart rate, and sleep patterns) via Bluetooth to the original customized apps and were further connected to the lifestyle observation platform via the OAuth 2.0 protocol. Environmental data in the patient's living environment (ie, temperature, humidity, and PM2.5 levels) were collected from the home air quality–sensing device and open environmental application programming interface (API). We also developed a health self-management smartphone app for patients and a lifestyle observation platform for medical staff, which together facilitate the continuous monitoring of lifestyle data and instant care advice. To assist physicians in organizing key information more effectively, data are visualized by combining trend charts, as shown in Figure 2. In addition to these trend charts, the daily prediction results are also used as a reference for decision support to help physicians better understand the status of their patients.

Figure 1. System architecture of the acute exacerbation of chronic obstructive pulmonary disease (AECOPD) prediction system. API: application programming interface; COPD: chronic obstructive pulmonary disease; mMRC: modified Medical Research Council dyspnea scale.

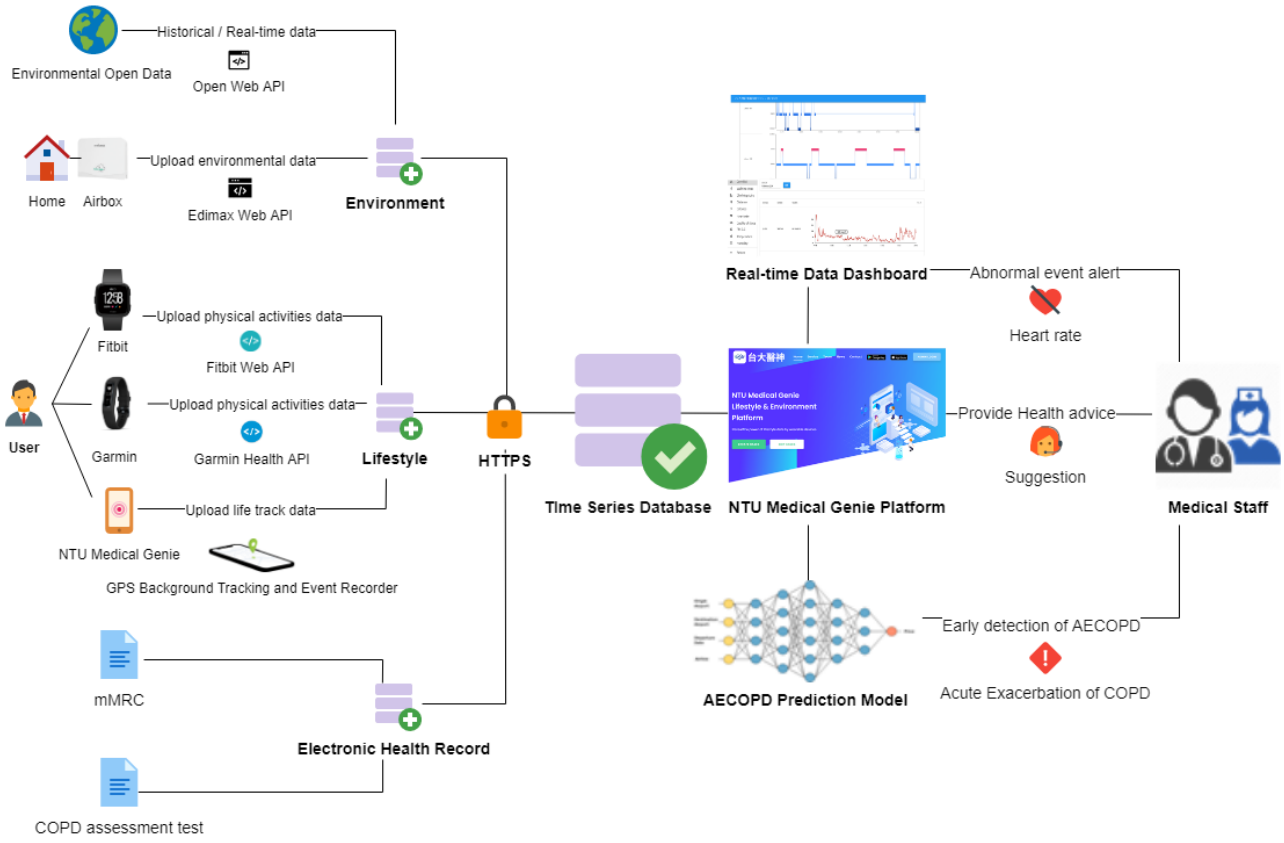
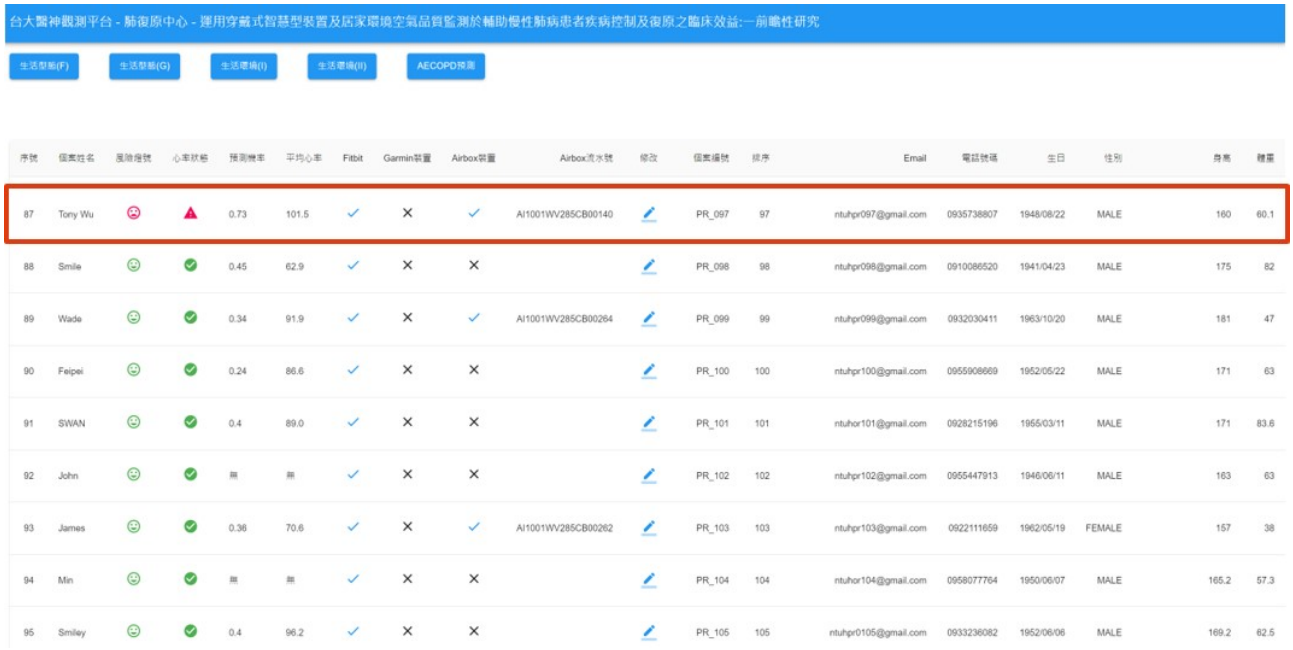


Figure 2. Data visualization from the lifestyle observation platform. Pm2.5: fine particulate matter.

Figure 3 shows an example system scenario: when the prediction probability exceeds 0.7, a red icon is displayed to prompt the case manager to intervene and take care of the patient. To protect patient privacy, the system transmits data via the HTTPS protocol, and personal information is encrypted. Figure 4 shows the data management workflow. Only verified medical service providers can access their patients' information, which ensures the confidentiality and integrity of the data. The mobile app was designed to record symptoms and produce trend charts to help patients better understand their health status, as shown in Figure 5. With appropriate expert advice, they are thus able to better manage their own health.

Figure 3. Daily prediction of acute exacerbation of chronic obstructive pulmonary disease.



序號	個案姓名	風險燈號	心率狀態	預測機率	平均心率	Fitbit	Garmin裝置	Airbox裝置	Airbox流水號	修改	個案編號	順序	Email	電話號碼	生日	性別	身高	體重
87	Tony Wu	☹	▲	0.73	101.5	✓	✗	✓	AI1001WV285CB00140		PR_097	97	ntuhpr097@gmail.com	0935738807	1948/08/22	MALE	160	60.1
88	Smile	☺	●	0.45	62.9	✓	✗	✗			PR_098	98	ntuhpr098@gmail.com	0910086520	1941/04/23	MALE	175	82
89	Wade	☺	●	0.34	91.9	✓	✗	✓	AI1001WV285CB00264		PR_099	99	ntuhpr099@gmail.com	0932030411	1963/10/20	MALE	181	47
90	Feipei	☺	●	0.24	86.6	✓	✗	✗			PR_100	100	ntuhpr100@gmail.com	0955908669	1952/05/22	MALE	171	63
91	SWAN	☺	●	0.4	89.0	✓	✗	✗			PR_101	101	ntuhor101@gmail.com	0928215196	1965/03/11	MALE	171	83.6
92	John	☺	●	無	無	✓	✗	✗			PR_102	102	ntuhpr102@gmail.com	0955447913	1946/06/11	MALE	163	63
93	James	☺	●	0.36	70.6	✓	✗	✓	AI1001WV285CB00262		PR_103	103	ntuhpr103@gmail.com	0922111659	1962/05/19	FEMALE	157	38
94	Min	☺	●	無	無	✓	✗	✗			PR_104	104	ntuhor104@gmail.com	0958077764	1950/06/07	MALE	165.2	57.3
95	Smiley	☺	●	0.4	96.2	✓	✗	✗			PR_105	105	ntuhpr0105@gmail.com	0933236082	1952/06/06	MALE	169.2	62.5

Figure 4. Data hierarchy workflow. PI: principal investigator.

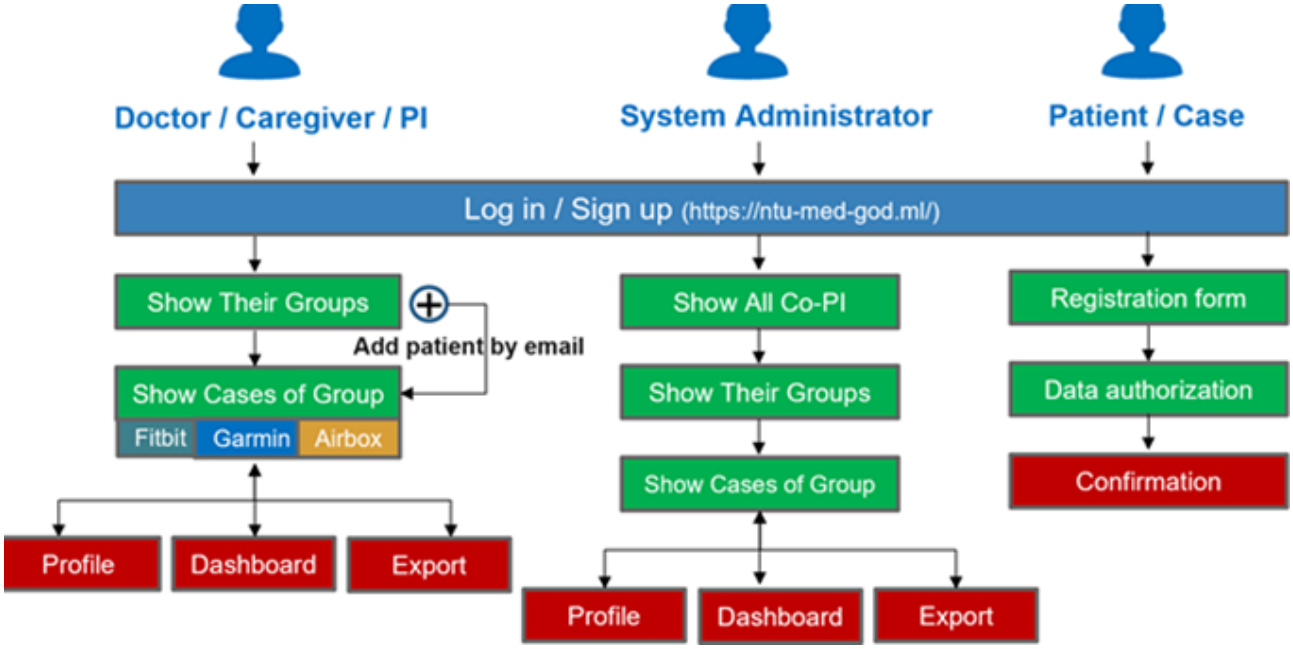


Figure 5. Screenshots of NTU-med-good health advice app.

Data Processing

The combination of clinical questionnaire data, environmental data, and physiological data was the main dataset used in our training model. To explore the influence of changes in environmental and lifestyle factors on AECOPD, first- and second-order differentiation models were applied to the environmental and physiological data to understand trends and serve as additional input features. Random downsampling was used to account for data imbalance, resulting in 5600 data points, one-third of which were used as the validation set and the remainder were used as the training set. We used forward-filling to pad missing and questionnaire data, as illustrated in Figure 6. The complete data selection rules are shown in Figure 7. Note that missing values is a common problem in data mining. A correlation analysis (Figure 8) was performed between physiological and environmental features to ensure that their interactions did not affect the prediction results. Finally, as shown in Textbox 1, 45 features were selected to predict the probability of AECOPD.

Figure 6. Forward padding for questionnaire data. AE: acute exacerbation.

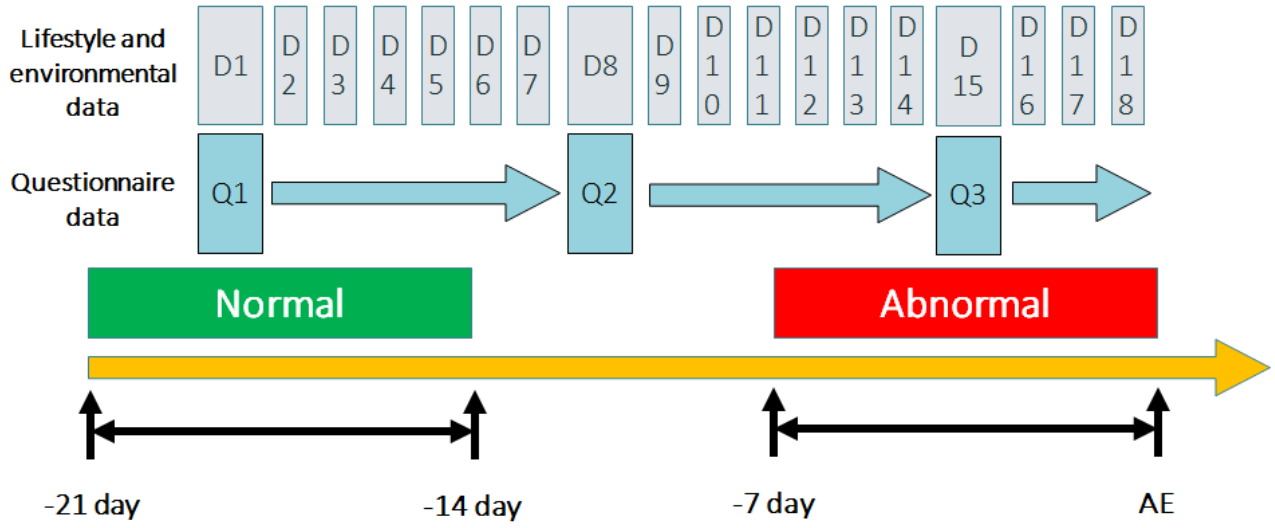


Figure 7. Decision rules for data selection. COPD: chronic obstructive pulmonary disease; AE: acute exacerbation.

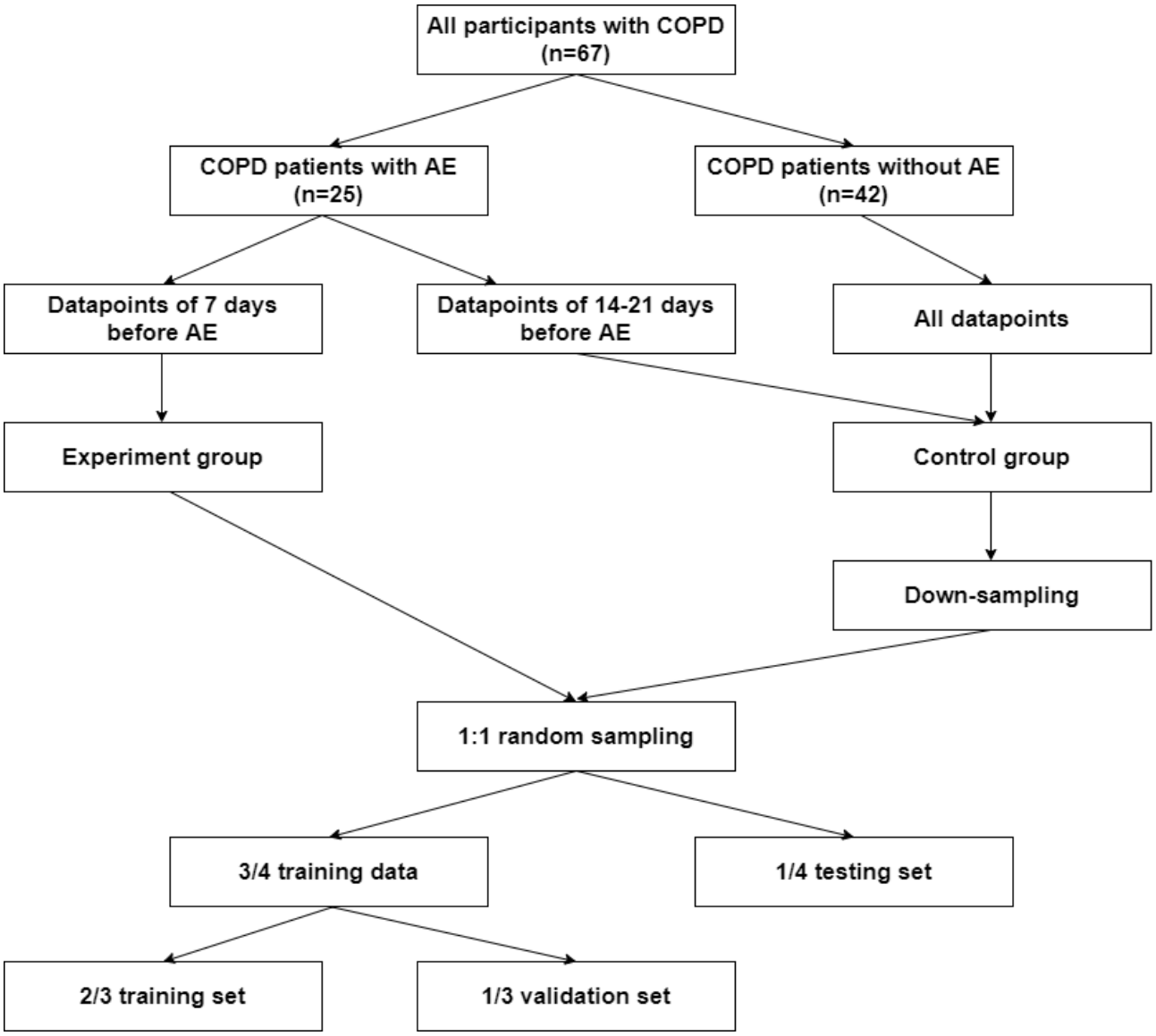
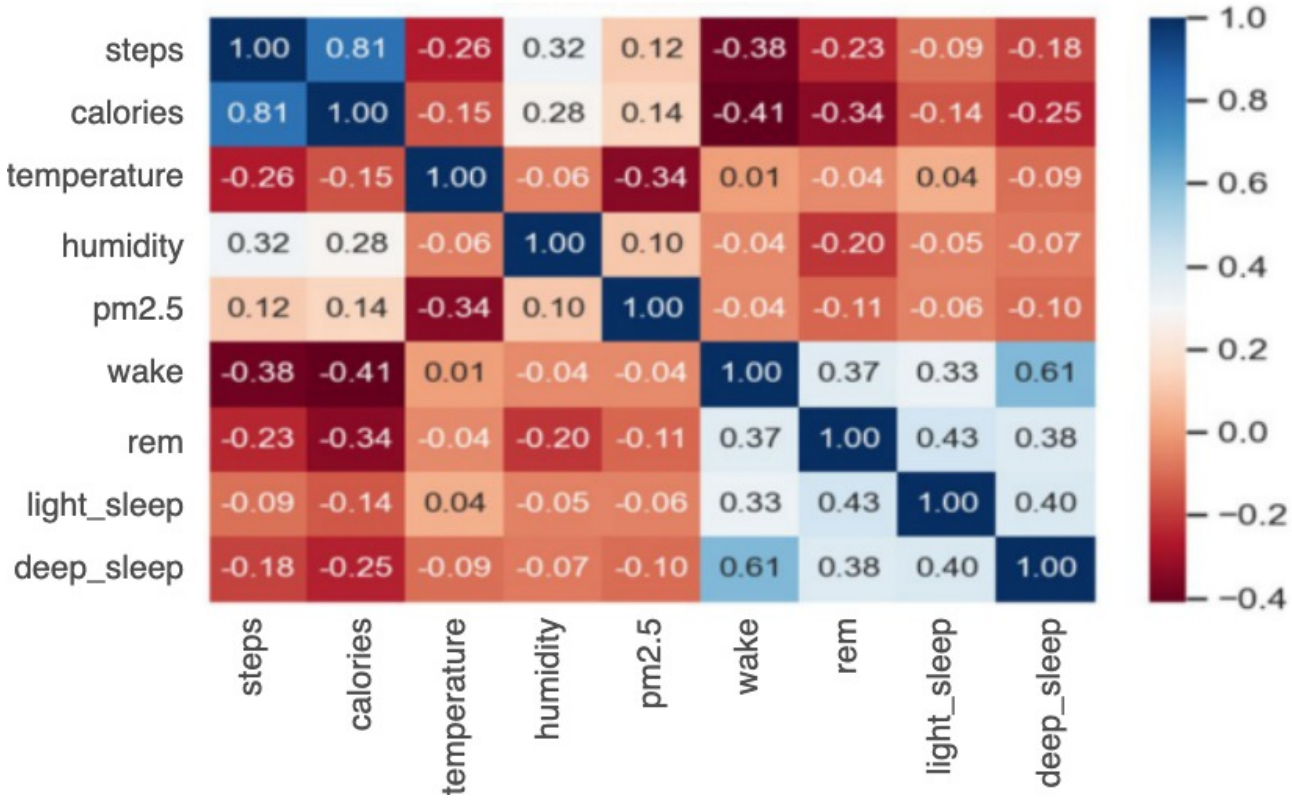


Figure 8. Correlation matrix of physiological and environmental data.



Textbox 1. Input data features of machine-learning and deep-learning models.

- Environmental factors

Temperature, humidity, fine particulate matter, first-order differentiation (diff1)_temperature, diff1_humidity, diff1_fine particulate matter, second-order differentiation (diff2)_temperature, diff2_humidity, diff2_fine particulate

- Physiological factors

Heart rate, walking steps, calories consumption, deep sleep time, light sleep time, rapid eye movement time, awake time, diff1_heart rate, diff1_ walking steps, diff1_calories consumption, diff1_deep sleep time, diff1_light sleep time, diff1_rapid eye movement time, diff1_awake time, diff2_heart rate, diff2_ walking steps, diff2_calories consumption, diff2_deep sleep time, diff2_light sleep time, diff2_rapid eye movement time, diff2_awake time

- Clinical questionnaires

Chronic obstructive pulmonary disease (COPD) assessment test (9 answers), modified Medical Research Council (mMRC) dyspnea scale (1 answer), life quality questionnaire (5 answers)

Classification Models

Classification algorithms for this study were selected according to previously published studies on COPD such as those of Wang et al [13] and Rahman et al [14]. The former group developed AECOPD identification models to reduce patient mortality and financial burdens, and the latter group attempted to identify relations between discriminatory heart rate variability features and disease severity in patients with pulmonary diseases and COPD. For model comparison with machine learning–based classification, we selected the following classifiers: decision trees [15], random forests [16], k-nearest neighbor clustering [17], linear discriminant analysis, and adaptive boosting [18]. We also propose a deep neural network (DNN) architecture for use in comparing the performance between machine-learning and deep-learning approaches on AECOPD prediction. Supervised learning was performed using AECOPD events and 51 features obtained from the lifestyle observation platform. Models were implemented using python libraries such as scikit-learn and Pytorch.

DNN Classification

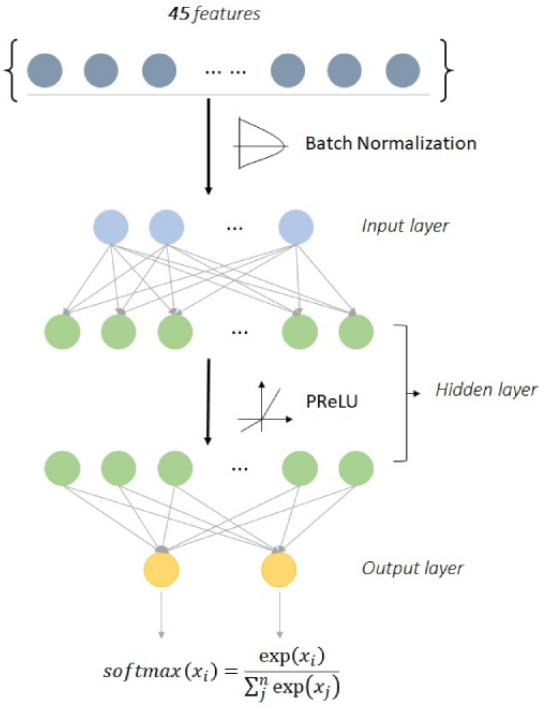
The DNN classification model was constructed using fully connected layers, which connect each neuron in one layer to every neuron in another layer, mapping feature representations to the target vector space. Hyperparameters for the two fully connected layers are presented in Table 1. Batch normalization was applied to input data sequences to reduce the internal covariate shift and gradient dependence [19]. For the activation function we used parametric rectified linear unit (PReLU), which combines the characteristics of ReLU and leaky ReLU, with the introduction of a variable slope α, randomly selected from

the uniform distribution during the training process to negative values [20]. The complete DNN architecture is shown in Figure 9.

Table 1. Hyperparameters of the machine-learning and deep-learning models.

Model and hyperparameters	Value
Decision trees	
min_samples_leaf	1
min_samples_split	2
AdaBoost: n_estimators	45
Random forests	
min_samples_leaf	1
min_samples_split	2
n_estimators	100
k-nearest neighbor	
n_neighbors	3
leaf_size	30
p	2
Deep neural network	
fully connected layer 1	45
fully connected layer 2	45

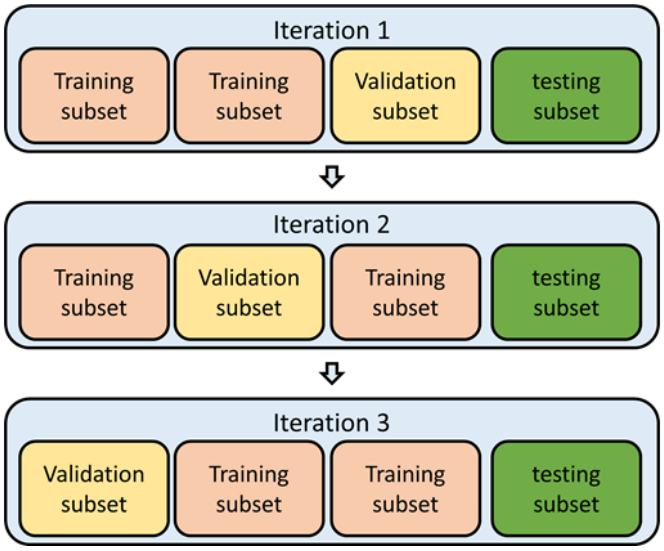
Figure 9. Deep neural network model architecture. PReLU: parametric rectified linear unit.



Validation and Model Assessment

We use 3-fold cross-validation to evaluate the stability of the prediction models. The workflow is shown in Figure 10. We used two metrics to evaluate the performance of the identification models based on the test set: the area under the receiver operating characteristic curve (AUROC) and the F1 score. We also used sensitivity, specificity, precision, and accuracy as assessment metrics.

Figure 10. Workflow of 3-fold cross-validation.



Results

Patient Characteristics

A total of 67 patients were registered for this study. Most of the patients were middle-aged (mean 66.62, SD 11.38 years) and were men. Eighteen percent had never smoked and the remainder were either current smokers or exsmokers. Detailed demographic information of the study participants is shown in Table 2.

Table 2. Demographics of study participants (N=67).

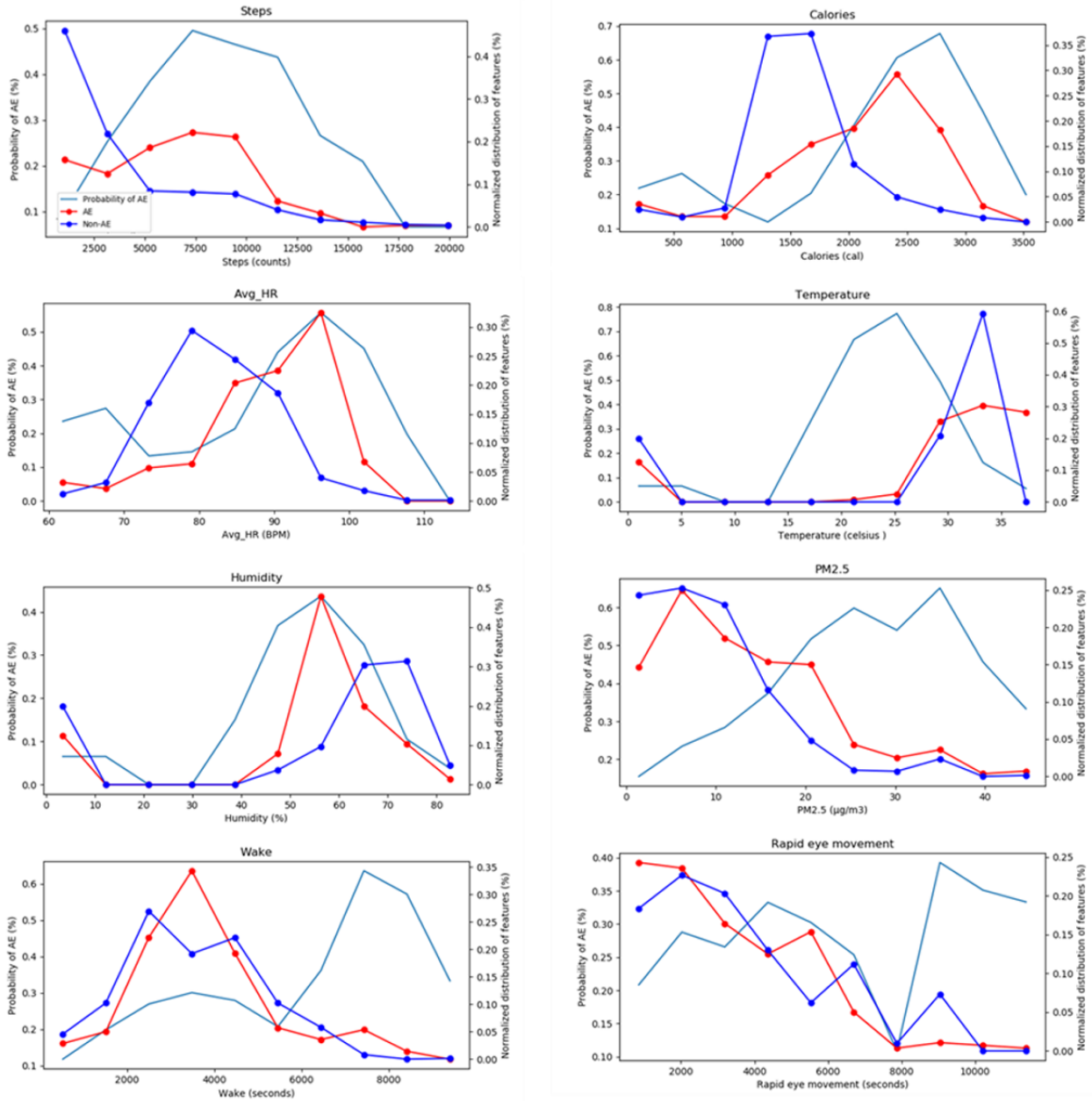
Characteristic	Value
Age (years), mean (SD)	66.62 (11.38)
Gender, n (%)	
Male	59 (88)
Female	8 (12)
Smoking history, n (%)	
Never smoker	18 (27)
Current smoker	9 (13)
Exsmoker	40 (60)
Comorbidities	
Diabetes mellitus	12 (18)
Hypertension	25 (37)
Myocardial infarction	1 (1)
Heart failure	2 (3)
Peripheral vascular disease	11 (16)
Bronchiectasis	15 (22)
Postnasal drip syndrome	6 (9)
Nasal septum deviation	5 (7)
Allergic rhinitis	19 (28)
Others	24 (36)
FEV1[a](% predicted), n (%)	
≥80	14 (21)
50-79	24 (36)
30-49	20 (30)
<30	9 (13)

[a]FEV1: postbronchodilator forced expiratory volume in 1 second.

Distribution of Physiological and Environmental Factors

Figure 11 illustrates the AECOPD probabilities versus the distributions of physiological and environmental features, among which average heart rate, PM2.5, steps walked, and calorie consumption were significantly different between those with and without AECOPD. This shows that physiological and environmental factors are useful for predicting AECOPD.

Figure 11. Acute exacerbation of chronic obstructive pulmonary disorder probability trends versus normalized distributions of physiological and environmental factors. HR: heart rate; AE: acute exacerbation: PM2.5: fine particulate matter.



AECOPD Prediction Model

Table 3 and Figure 12 demonstrate the performance of the implemented models. The DNN model yielded the best performance with 6 metrics higher than 90%.

To determine which model best fits diverse scenarios, we trained the model using various combinations of data features, as shown in Table 4: the prediction including all of the features yielded the best performance. These results further confirmed that physiological and environmental data features are more predictive of AECOPD than conventional clinical questionnaires.

Table 3. Performance of each model with all features input.

Model	Accuracy	AUCROC[a]	Sensitivity	Specificity	Precision	F1
Random forests	0.914	0.986	0.877	0.955	0.955	0.914
Decision trees	0.792	0.796	0.712	0.881	0.867	0.782
kNN[b]	0.743	0.779	0.712	0.776	0.776	0.743
LDA[c]	0.829	0.882	0.781	0.881	0.877	0.826
AdaBoost[d]	0.886	0.969	0.822	0.955	0.952	0.882
DNN[e]	0.921	0.964	0.904	0.940	0.943	0.923

[a]AUROC: area under the receiver operating characteristic curve.

[b]kNN: k-nearest neighbor.

[c]LDA: linear discriminant analysis.

[d]AdaBoost: adaptive boosting.

[e]DNN: deep neural network.

Figure 12. Receiver operating characteristic (ROC) curve and area under the receiver operating characteristic curve of all models. AE: acute exacerbation; KNN: k-nearest neighbor; LDA: linear discriminant analysis.

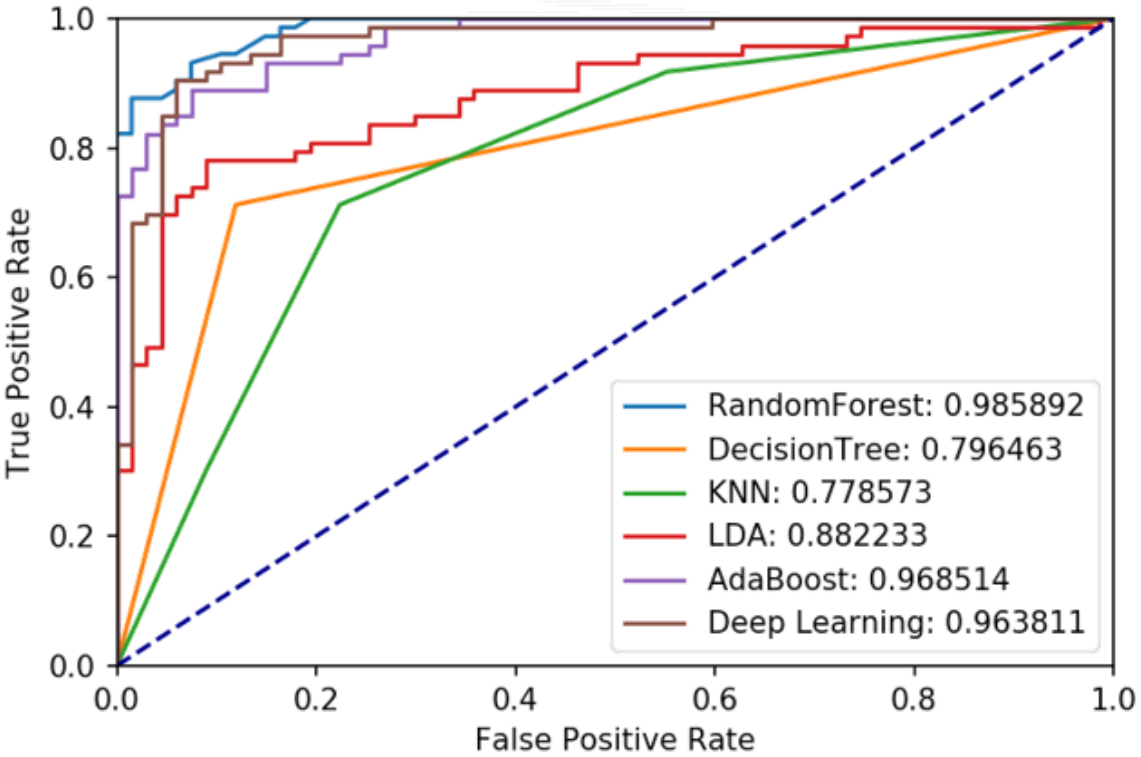


Table 4. Performance given different feature sets.

Feature	Model	Accuracy	AUROC[a]	Sensitivity	Specificity	Precision	F1
All features	DNN[b]	0.9357	0.9699	0.9452	0.9253	0.9393	0.9323
Lifestyle	Random forests	0.8428	0.9195	0.9253	0.7671	0.9180	0.8358
Env[c]	Decision trees	0.8000	0.8185	0.8805	0.7260	0.8688	0.7910
Env+Lifestyle	Random forests	0.8357	0.9256	0.8805	0.7945	0.8787	0.8345
Clinical questionnaire	AdaBoost[d]	0.6956	0.6825	0.6666	0.7142	0.7692	0.7407

[a]AUROC: area under the receiver operating characteristic curve.

[b]DNN: deep neural network.

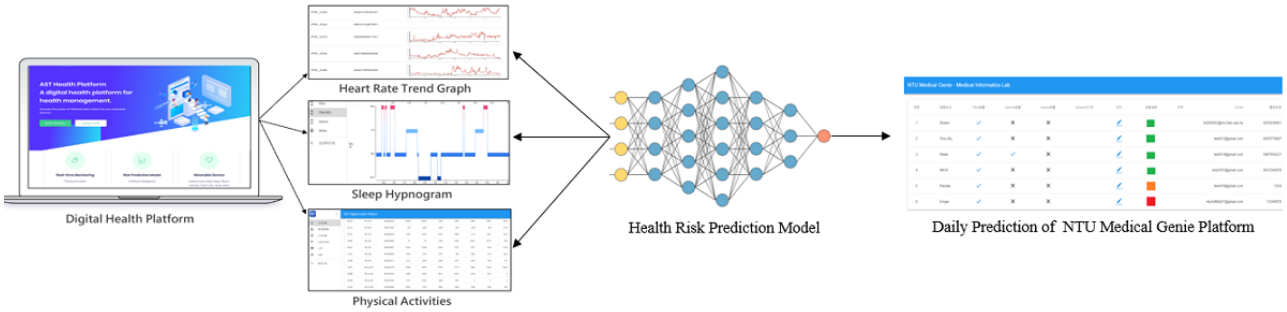
[c]Env: environmental.

[d]AdaBoost: adaptive boosting.

AECOPD Prediction System

To account for incomplete data, which is typical in real-world apps, the prediction system supports AECOPD prediction via optional features. When only lifestyle or environmental data are automatically uploaded daily, the system still predicts whether AECOPD will occur within the next 7 days. Therefore, multiple AECOPD prediction models were deployed on the server. Through the process shown in Figure 13, daily prediction results are provided to support physicians in making decisions. Different color signs displayed on the system indicate different risk levels.

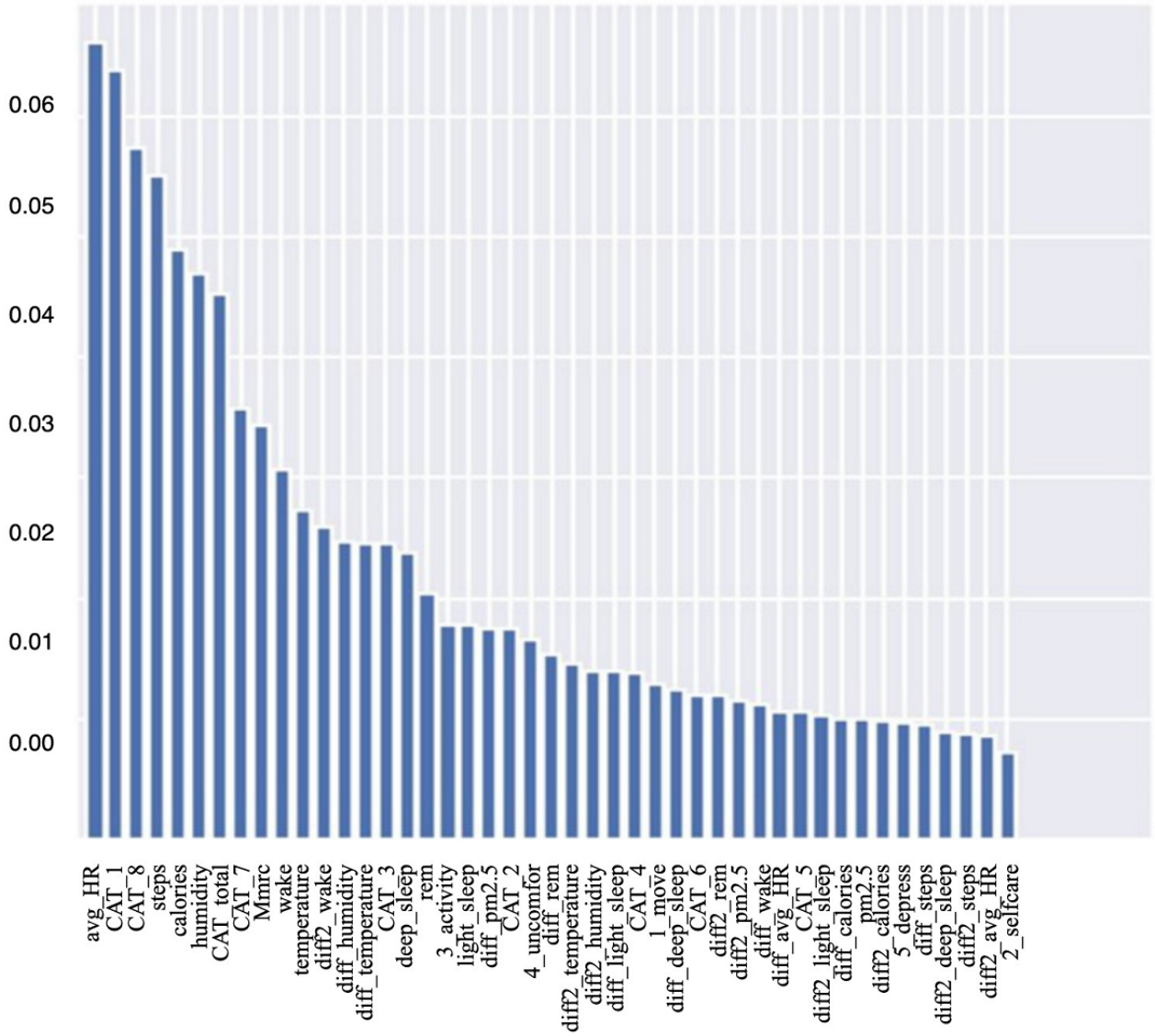
Figure 13. Acute exacerbation of chronic obstructive pulmonary disease prediction system.



Feature Importance

Figure 14 shows the importance scores of model features as evaluated by a random forest algorithm. Feature importance is a measure of the ability to improve the purity of the random forest model's leaf nodes. Daily activity–related features such as average heart rate, calorie consumption, and steps walked yielded higher importance scores, which indicates that these features have greater potential to improve the performance of the random forest model. As average heart rate had the highest importance score, and is thus likely the most influential predictor of AECOPD, we used a warning sign for abnormal heart rates in the AECOPD prediction system to alert physicians when the patient's average heart rate was abnormal.

Figure 14. Feature importance scores as evaluated by random forest.



Discussion

Principal Findings

We implemented an AECOPD prediction system by integrating wearable devices, Internet-of-Things environment sensors, a smartphone app, machine learning, and deep learning. We present six models for AECOPD identification. Additionally, we selected features to determine the optimal feature set for this task. The performance of each model is demonstrated according to sensitivity, specificity, F1-score, accuracy, precision, and AUROC metrics obtained based on 3-fold cross-validation. To the best of our knowledge, this is the most comprehensive study that used machine-learning models to predict AECOPDs.

Clinical questionnaires tend to be more subjective, which can affect clinical decisions. The AECOPD prediction model with all data features achieved the best performance. These results showed that physiological and environmental data are more powerful predictors than questionnaire data. Compared with clinical questionnaire data alone, lifestyle and environmental data yielded improvements of 10% in accuracy and 20% in AUROC.

Comparison With Prior Work

In the 2010s, researchers began to attempt to predict COPD exacerbation. One study used demographic features, vital signs, and electronic medical records to predict COPD exacerbations in the emergency department [21]. Another used 28 features, including vital signs, medical history, inflammatory indicators, and tree-based machine learning, to predict the prognosis of hospitalized patients with COPD [22]. In contrast, in this study, we focused on exacerbation risk prediction for discharged COPD patients, because their health condition is likely to be less accessible. Another study remotely monitored AECOPD in patients via questionnaire data. They demonstrated an accuracy of 100% for event-based prediction and up to 80.5% for symptom-based prediction [23]. In addition, Shah et al [24] used pulse oximetry and three vital signs to predict AECOPD, reaching a mean AUROC of 68%. In comparison to these studies, we used daily activities and environmental information as predictors to trace the health conditions of patients with COPD and achieved higher performance. With wearable devices and smartphone apps, all relevant COPD information can be collected instantly. Such a system will be helpful for achieving the goal of personalized health management in the future. Thus,

overall, our study constitutes a novel solution making use of various data sources for superior AECOPD prediction performance.

Limitations

Because of limitations in the air quality–sensing device, environmental data collection was restricted to the user's bedroom, which degrades the prediction results. To account for this, in the future we plan to use GPS functions to trace the user's movements to capture the corresponding environmental data published via governmental open APIs.

In contrast to assumptions of physicians, Figure 10 shows that patients with AECOPD engaged in more physical activity than those without AECOPD. After reviewing the personal lifestyle of each patient, we found that some engaged in intense exercise even if they were uncomfortable, which goes against the accepted knowledge of the medical profession. In the future, more data could help to shed light on this apparent paradox.

Conclusions

Patients with COPD generally must return to the hospital monthly for numerous clinical tests, which is a time-consuming procedure. At the same time, it is impractical for discharged patients or those under home care to continuously observe their health conditions. They thus run the risk of AECOPDs between routine visits.

In this study, we attempted to predict whether a patient with COPD will experience acute exacerbation of their condition within the next 7 days. In general, lifestyle and environmental data of patients are difficult to collect effectively. However, with the proposed system, all COPD-related data are uploaded automatically. Our results indicate that lifestyle and environmental data facilitate the precise management of users' health conditions, and can even produce early warnings of AECOPD. The experimental results confirmed that such lifestyle and environmental data are highly correlated to user health conditions. In the future, we will enhance the prediction system and perform external validation to ensure that the model can be applied to other regions.

Acknowledgments

This research is supported by the Ministry of Science and Technology, Taiwan (grant MOST 110-2634-F-002-032-).

Conflicts of Interest

None declared.

References

1. Qureshi H, Sharafkhaneh A, Hanania NA. Chronic obstructive pulmonary disease exacerbations: latest evidence and clinical implications. Ther Adv Chronic Dis 2014 Sep 14;5(5):212-227 [FREE Full text] [doi: 10.1177/2040622314532862] [Medline: 25177479]
2. Yan R, Wang Y, Bo J, Li W. Healthy lifestyle behaviors among individuals with chronic obstructive pulmonary disease in urban and rural communities in China: a large community-based epidemiological study. Int J Chron Obstruct Pulmon Dis 2017;12:3311-3321. [doi: 10.2147/COPD.S144978] [Medline: 29180861]
3. Global Initiative for Chronic Obstructive Lung Disease. URL: https://goldcopd.org/ [accessed 2021-03-23]
4. Ambrosino N, Bertella E. Lifestyle interventions in prevention and comprehensive management of COPD. Breathe (Sheff) 2018 Sep 31;14(3):186-194 [FREE Full text] [doi: 10.1183/20734735.018618] [Medline: 30186516]
5. Adibi A, Sin DD, Safari A, Johnson KM, Aaron SD, FitzGerald JM, et al. The Acute COPD Exacerbation Prediction Tool (ACCEPT): a modelling study. Lancet Respir Med 2020 Oct;8(10):1013-1021. [doi: 10.1016/S2213-2600(19)30397-2] [Medline: 32178776]
6. Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review. Chron Respir Dis 2016 Aug;13(3):264-283 [FREE Full text] [doi: 10.1177/1479972316642365] [Medline: 27097638]
7. Hurst JR, Donaldson GC, Quint JK, Goldring JJ, Patel AR, Wedzicha JA. Domiciliary pulse-oximetry at exacerbation of chronic obstructive pulmonary disease: prospective pilot study. BMC Pulm Med 2010 Oct 20;10:52 [FREE Full text] [doi: 10.1186/1471-2466-10-52] [Medline: 20961450]
8. Sánchez-Morillo D, Crespo M, León A, Crespo Foix LF. A novel multimodal tool for telemonitoring patients with COPD. Inform Health Soc Care 2015 Jan 31;40(1):1-22. [doi: 10.3109/17538157.2013.872114] [Medline: 24380372]
9. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012 Jul 26;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]
10. Saleh L, Mcheick H, Ajami H, Mili H, Dargham J. Comparison of machine learning algorithms to increase prediction accuracy of COPD domain. In: Mokhtari M, Abdulrazak B, Aloulou H, editors. Enhanced Quality of Life and Smart Living. ICOST 2017. Lecture Notes in Computer Science. Cham: Springer; 2017:247-254.

11. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Intern Med 1987 Feb;106(2):196-204. [doi: 10.7326/0003-4819-106-2-196] [Medline: 3492164]
12. Riis HC, Jensen MH, Cichosz SL, Hejlesen OK. Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. J Med Eng Technol 2016;40(1):1-7. [doi: 10.3109/03091902.2015.1105317] [Medline: 26745746]
13. Wang C, Chen X, Du L, Zhan Q, Yang T, Fang Z. Comparison of machine learning algorithms for the identification of acute exacerbations in chronic obstructive pulmonary disease. Comput Methods Programs Biomed 2020 May;188:105267. [doi: 10.1016/j.cmpb.2019.105267] [Medline: 31841787]
14. Rahman MJ, Nemati E, Rahman MM, Nathan V, Vatanparvar K, Kuang J. Automated assessment of pulmonary patients using heart rate variability from everyday wearables. Smart Health 2020 Mar;15:100081. [doi: 10.1016/j.smhl.2019.100081]
15. Himani S, Sunil K. A survey on decision tree algorithms of classification in data mining. Int J Sci Res 2016 Apr 05;5(4):2094-2097 [FREE Full text] [doi: 10.21275/v5i4.nov162954]
16. Ho TK. A data complexity analysis of comparative advantages of decision forest constructors. Pattern Anal Appl 2002 Jun 7;5(2):102-112. [doi: 10.1007/s100440200009]
17. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Amer Statist 1992 Aug;46(3):175-185. [doi: 10.1080/00031305.1992.10475879]
18. Cristóbal E, Javier M, Fernando S, Myriam A, Amaia A, Begona G, et al. Machine learning for COPD exacerbation prediction. Eur Resp J 2015;46(Suppl 59):OA3282. [doi: 10.1183/13993003.congress-2015.OA3282]
19. Sergey I, Christian S. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv. 2015. URL: https://arxiv.org/abs/1502.03167 [accessed 2021-04-11]
20. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 Presented at: IEEE International Conference on Computer Vision; 2015; Santiago, Chile URL: https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/He_Delving_Deep_into_ICCV_2015_paper.pdf
21. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med 2018 Sep;36(9):1650-1654. [doi: 10.1016/j.ajem.2018.06.062] [Medline: 29970272]
22. Peng J, Chen C, Zhou M, Xie X, Zhou Y, Luo CH. A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. Sci Rep 2020 Feb 20;10(1):3118. [doi: 10.1038/s41598-020-60042-1] [Medline: 32080330]
23. Fernandez-Granero MA, Sanchez-Morillo D, Lopez-Gordo MA, Leon A. A machine learning approach to prediction of exacerbations of chronic obstructive pulmonary disease. In: Artificial Computation in Biology and Medicine. Lecture Notes in Computer Science Volume 9107. Cham: Springer; 2015 Presented at: International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC 2015); June 1-5, 2015; Elche, Spain p. 305-311. [doi: 10.1007/978-3-319-18914-7_32]
24. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]

Abbreviations

AECOPD: acute exacerbation of chronic obstructive pulmonary disease
API: application programming interface
AUROC: area under the receiver operating characteristic curve
CAT: chronic obstructive pulmonary disease assessment test
COPD: chronic obstructive pulmonary disease
DNN: deep neural network
GOLD: Global initiative for chronic obstructive lung disease
mMRC: modified Medical Research Council dyspnea scale
PM2.5: fine particulate matter

Edited by L Buis; submitted 17.07.20; peer-reviewed by D Ruttens, X Li; comments to author 13.10.20; revised version received 30.01.21; accepted 23.03.21; published 06.05.21.

Please cite as:
Wu CT, Li GH, Huang CT, Cheng YC, Chen CH, Chien JY, Kuo PH, Kuo LC, Lai F
Acute Exacerbation of a Chronic Obstructive Pulmonary Disease Prediction System Using Wearable Device Data, Machine Learning, and Deep Learning: Development and Cohort Study
JMIR Mhealth Uhealth 2021;9(5):e22591
URL: https://mhealth.jmir.org/2021/5/e22591
doi:10.2196/22591
PMID:33955840

Original Paper

Mobile Health–Based Thermometer for Monitoring Wound Healing After Endovascular Therapy in Patients With Chronic Foot Ulcer: Prospective Cohort Study

Donna Shu-Han Lin[1,2], MD; Jen-Kuang Lee[1,2,3,4,5], MD, PhD

[1]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[2]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan

[3]Department of Laboratory Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

[4]Telehealth Center, National Taiwan University Hospital, Taipei, Taiwan

[5]Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Jen-Kuang Lee, MD, PhD
Division of Cardiology
Department of Internal Medicine
National Taiwan University Hospital
No 7 Rd. South Chung Shan
Taipei, 100
Taiwan
Phone: 886 223123456 ext 65412
Email: b85401104@gmail.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2022/6/e39749/

Abstract

Background: Foot temperature may increase after endovascular therapy, but the relationship between foot temperature and wound healing is unclear.

Objective: This study was performed to evaluate the feasibility of a mobile health (mHealth)–based thermometer for foot temperature monitoring in patients with chronic foot ulcer before and after endovascular therapy and to determine the association between temperature change and wound healing time.

Methods: This was a prospective cohort study. Patients who had a chronic foot ulcer (>3 months) and underwent endovascular therapy between July 2019 and December 2019 were included. The participants received standard medical care and endovascular therapy for revascularization. The mHealth-based thermometer, composed of 4 temperature-sensing chips, was put on the foot before and after endovascular therapy. Data from the chips were transferred to an associated mobile phone app via Bluetooth. Wound healing time was estimated using the Kaplan-Meier method, and the associations between baseline characteristics and clinical outcomes were evaluated using a Cox proportional hazard model.

Results: A total of 163 patients with chronic foot ulcer who underwent endovascular therapy were enrolled and followed up until wound healing was complete or for 180 days. The mean foot temperature before endovascular therapy was 30.6 (SD 2.8 °C). Foot temperature increased significantly (mean 32.1 °C, SD 2.8 °C; *P*=.01) after the procedure. Wound healing time was significantly different in the Kaplan-Meier curves of the patient group with temperature changes ≥2 °C and the group with temperature changes ≤2 °C (log-rank *P*<.001). A foot temperature increase ≥2 °C after endovascular therapy was associated with increased wound healing in univariate analysis (hazard ratio [HR] 1.78, 95% CI 1.24-2.76, *P*=.02), and the association remained significant in multivariate analysis (HR 1.69, 95% CI 1.21-2.67, *P*=.03).

Conclusions: The mHealth-based thermometer was feasible and useful for foot temperature monitoring, which may provide health care professionals with a new endpoint for endovascular therapy. Foot temperature increases ≥2 °C after endovascular

therapy were associated with faster wound healing in patients with chronic foot ulcer. Further studies are needed, however, to confirm these findings.

(JMIR Mhealth Uhealth 2021;9(5):e26468) doi:10.2196/26468

KEYWORDS

temperature; peripheral artery disease; endovascular therapy; mHealth; app; foot; therapy; wound; thermometer; monitoring; ulcer; artery; prospective; cohort; healing

Introduction

Chronic foot ulcer is a disease associated with multiple conditions [1], such as old age, diabetes, smoking, and chronic kidney disease. It often occurs in patients with the most extreme cases of lower extremity arterial disease and inflammatory cardiac disease [2,3]. Patients presenting with minor or major tissue loss (Rutherford stages V and VI) are at high risk of future amputation [4]. Current guidelines recommend reperfusion with bypass surgery or endovascular therapy for patients with critical limb ischemia to minimize tissue loss [5]. Optimal endpoints in endovascular therapy are important to determine wound recovery in order to prevent future amputation.

Although the guidelines [5,6] suggest the use of an in-line revascularization strategy for endovascular therapy, various procedural techniques and outcomes have been advocated. Iida et al [7] reported that angiosome-guided endovascular therapy improved wound healing rates compared to the wound healing rates of non–angiosome-guided procedures, emphasizing the concept of anatomy. Kawarada et al [8] found that skin perfusion pressure, a physiology-based parameter, was useful in determining when to end endovascular therapy. Moreover, transcutaneous oxygen pressure, near-infrared spectroscopy, wound blushing phenomenon, and vascular flow reserve have all been studied as endovascular therapy outcome measures [9-14]. It has been reported that periwound foot temperature is associated with wound status [15]. Foot temperature rises immediately after endovascular therapy, but the relationship between foot temperature and wound healing in endovascular therapy has never been studied.

The use of traditional thermometers in measuring foot temperature is limited by their poor sensitivity. With the newest technology, however, a nanochip thermometer can be used to measure temperature in various places—solid objects, air, skin, etc—easily, rapidly, and accurately. Connecting nanotechnology to mobile health (mHealth) is currently a trend for health monitoring and diagnosis to improve health care and optimize individualized medicine. With mHealth, patients can operate, record, and upload medical information to cyberspace through mobile apps independently, that is, without clinic visits or doctor consultations.

We aimed to investigate the feasibility of using an mHealth-based thermometer to monitor foot temperature in patients with chronic foot ulcer and lower extremity arterial disease before and after endovascular therapy. In addition, we examined whether increases in foot temperature were associated with wound healing times in these patients.

Methods

Study Design

This study was a prospective cohort study performed at a single tertiary medical center in Taipei, Taiwan. The study was conducted with the approval of the institutional review board of National Taiwan University Hospital (201906078RIPD) and in accordance with the Declaration of Helsinki [16]. All participants provided written informed consent both for the procedure and for inclusion in the study cohort.

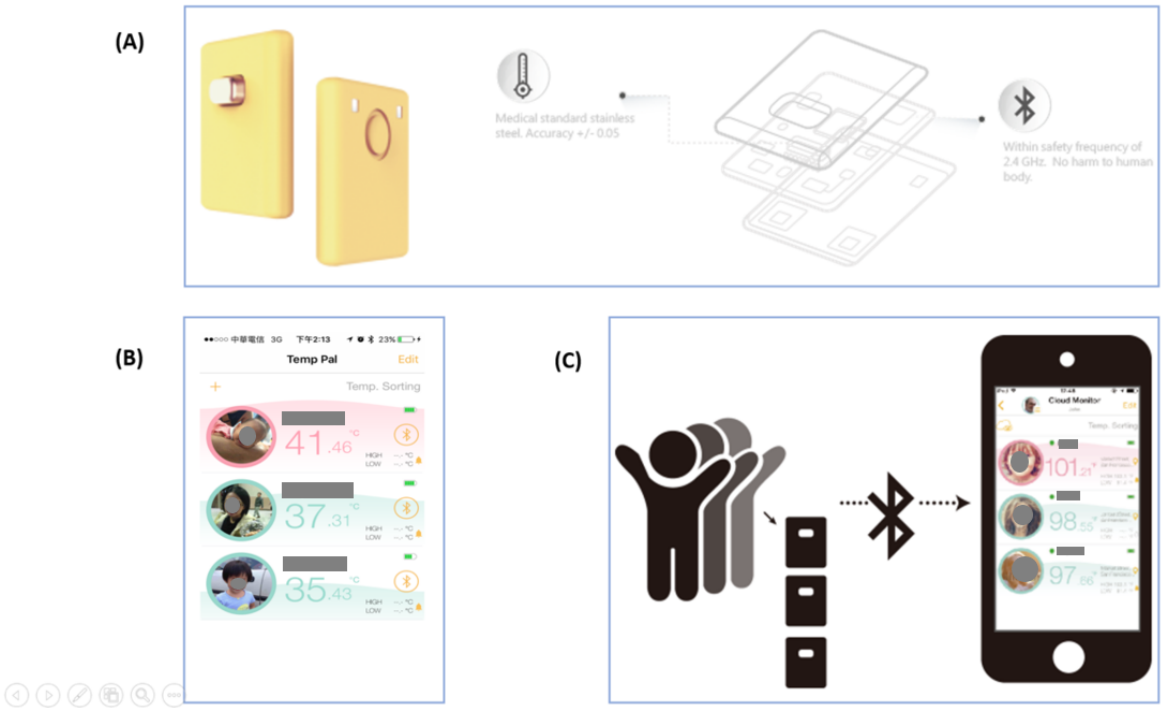
Patient Selection

Patients with chronic foot ulcers were consecutively enrolled from the outpatient clinic or during admission from July 2019 to December 2019 if they were adults scheduled to undergo endovascular therapy of the iliac artery, superficial femoral artery, popliteal artery, anterior tibial artery, peroneal artery, and posterior tibial artery. Demographic, clinical, and laboratory data were obtained prospectively, with a corroborative retrospective monthly review performed for all patients who underwent endovascular therapy through quality-control reports of the catheterization laboratory. The exclusion criteria for the study were patients with an active infection, with Rutherford stage VI ischemia, who did not ultimately receive endovascular therapy, and without recorded pre– or post–endovascular therapy foot temperature. Patients had follow-up visits at the outpatient department in our hospital at least every 3 months for wound healing assessment.

mHealth-Based Thermometer

Temp Pal (iWEECARE Co Ltd; Figure 1) is a wireless continuous body temperature monitoring device that allows simultaneous recordings at multiple spots. We used 4 devices paired with one iOS mobile device and the Temp Pal app via Bluetooth (Figure 1) to capture continuous temperature data, which were uploaded to the cloud and exported. With an auto-calibration algorithm, the accuracy and precision of the Temp Pal is within ±0.05 °C.

Figure 1. (A) Temp Pal body-temperature monitoring nanochip-based thermometer and app, which can (B) monitor multiple devices simultaneously and (C) transfer data via Bluetooth.



.

Outcome Measures

The primary endpoint was complete wound healing, which was defined as complete epithelialization of the wound determined with visual assessment performed by doctors. Patients were followed up for up to 180 days after endovascular therapy or until wound healing was complete, whichever occurred first. We used time-to-event analysis for all outcomes. Secondary outcomes were change in foot temperature at the instep of the foot and at the sole after endovascular therapy.

Statistical Analysis

Data were analyzed using SPSS software (version 22; IBM Corp). Statistical significance was considered to be indicated by 2-sided $P<.05$. Nonnormally distributed continuous data are reported as median and IQR. The chi-square test and Fisher exact test were used to compare differences in discrete and categorical variables (sex, diabetes, hypertension), respectively. The *t* test or the Wilcoxon rank-sum test was to compare continuous variables (foot temperature) pre– and post–endovascular therapy. The correlation between instep and sole temperature was examined using Pearson correlation.

The association between demographic, clinical, and laboratory characteristics with wound healing 180 days after endovascular therapy was determined with univariate logistic regression analysis. Variables achieving statistical significance ($P<.10$) in the univariate analysis were included in multivariate logistic regression analysis. Collinearity testing was done to avoid interdependence between the model variables. Hazard ratios (HRs) with 95% CIs are reported.

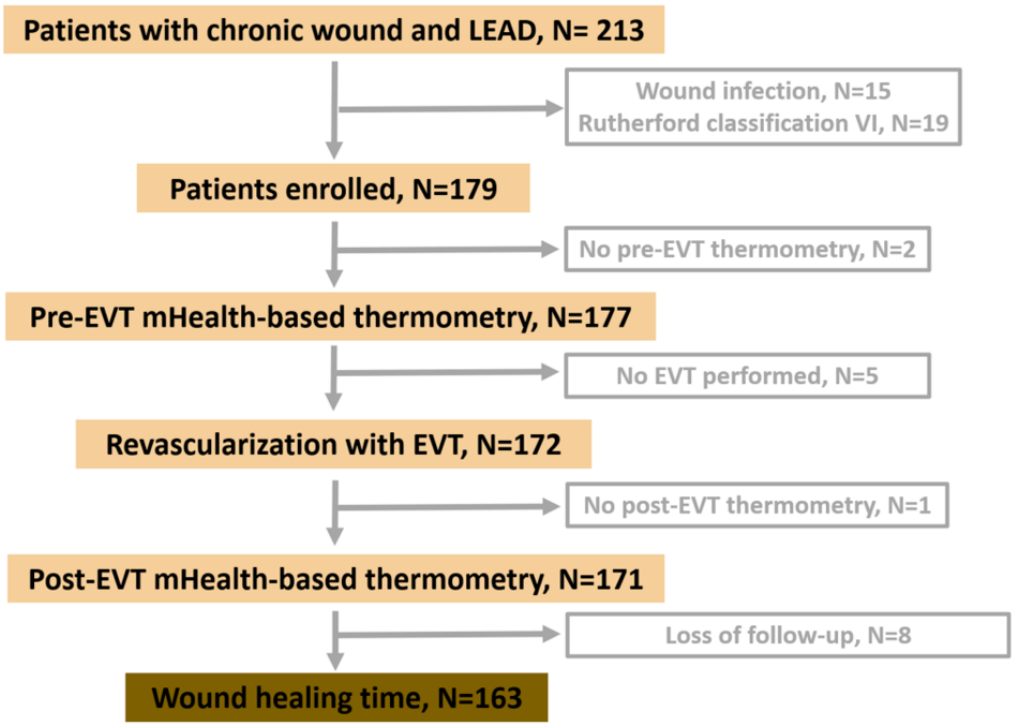
The threshold of foot temperature increase associated with wound healing was identified in the multivariate analysis. Time-to-wound healing was compared between patients whose foot temperature increase was greater than and those whose foot temperature increase was less than this threshold using the log-rank test or Gehan-Breslow-Wilcoxon Test. The Kaplan–Meier technique was used to estimate the cumulative incidence of complete wound healing in 180 days for each of the 2 groups.

Results

Patient Demographics and Clinical Features

Between July 2019 and December 2019, a total of 213 patients who had a chronic foot ulcer for longer than 3 months signed informed consent forms and were screened for enrollment (Figure 2). Of these, 15 patients were excluded for clinical signs of wound infection, and 19 patients were excluded due to Rutherford stage VI ischemia. Among the remaining 179 patients, 2 patients did not undergo pre–endovascular therapy foot temperature measurement, 5 patients did not undergo endovascular therapy, and 1 patient did not undergo post–endovascular therapy temperature measurement. Eight other patients were lost to follow-up during the study period; therefore, 163 patients were included in the analysis.

Figure 2. Participant flowchart. LEAD: lower extremity arterial disease; EVT: endovascular therapy.



The mean follow-up was 133 days (SD 55). The mean age of the patients was 72.6 years (SD 13.1), and 104 patients (63.8%) were male. The mean BMI was 24.3 kg/m^2 (SD 4.5). There were 50 (30.6%) patients who were wheelchair-bound or bedridden. Notable comorbidities included coronary artery disease (125/163, 76.7%), diabetes (123/163, 75.4%), hypertension (128/163, 78.5%), dyslipidemia (112/163, 68.7%), and dialysis dependence (37/163, 22.6%). The majority of patients were taking antiplatelet medication (140/163, 85.8%), and 138 (84.6%) patients were taking medication for diabetes (hemoglobin A_{1c} level: median 7.8%, IQR 2.7%). Baseline demographic and clinical characteristics are shown in Table 1.

Table 1. Baseline patient characteristics.

Characteristic	Value (n=163)
Age, mean (SD)	72.6 (13.1)
BMI[a], mean (SD)	24.3 (4.5)
Sex, n (%)	
Male	104 (63.8)
Female	59 (36.2)
Ambulatory, n (%)	
Yes	113 (69.3)
No	50 (30.7)
Risk factors, n (%)[b]	
Current smoking	44 (26.9)
Diabetes	123 (75.4)
Hypertension	128 (78.5)
Dyslipidemia	112 (68.7)
Chronic kidney disease	89 (54.6)
Dialysis	37 (22.6)
Atrial fibrillation	34 (21.0)
Coronary artery disease	125 (76.7)
Congestive heart failure	34 (20.8)
Stroke	20 (12.2)
Left ventricular ejection fraction (%), median (IQR)	55.2 (16.0)
Hemoglobin A_{1C} (%), median (IQR)	7.8 (2.7)
Medications, n (%)[c]	
Antiplatelet	140 (85.8)
Anticoagulation	34 (20.8)
Beta-blocker	63 (38.6)
Calcium channel blockers	86 (52.7)
Angiotensin converting enzyme inhibitors/angiotensin receptor blockers	94 (57.6)
Oral antidiabetic drugs	138 (84.6)
Insulin	34 (20.8)

[a]BMI: body mass index.

[b]Percentages do not add to 100 since patients may have multiple comorbidities.

[c]Percentages do not add to 100 since patients may take multiple medications.

Foot Temperature Before and After Endovascular Therapy

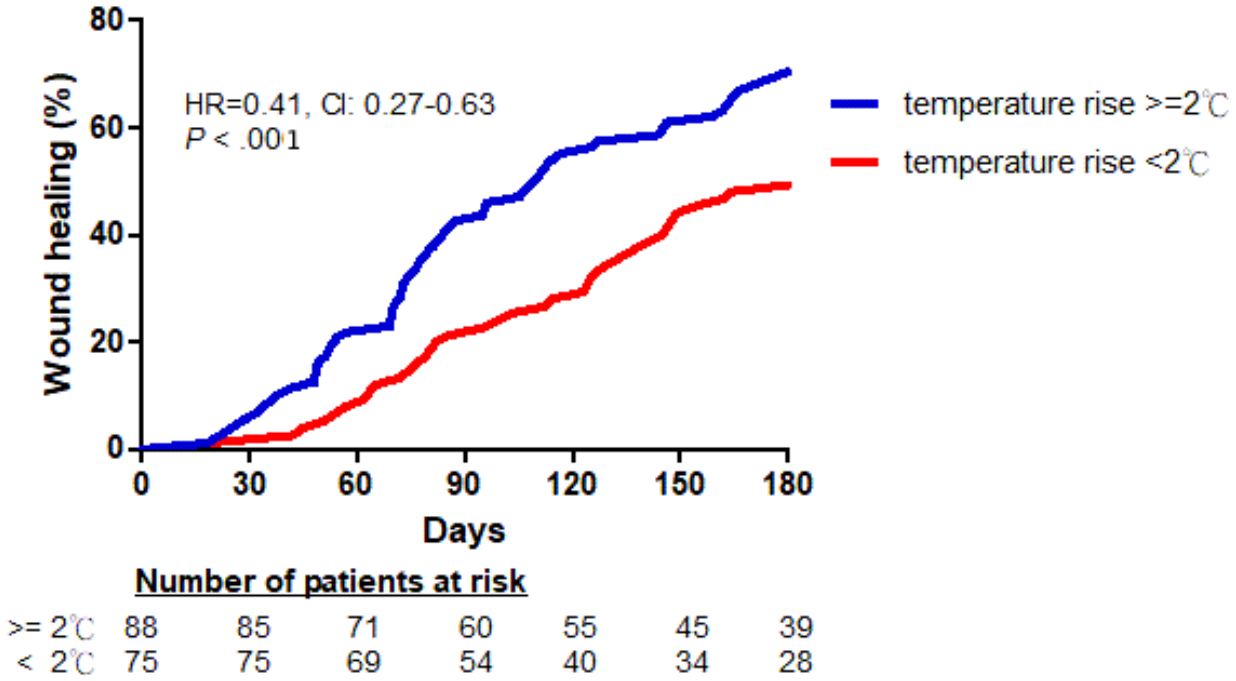
Instep temperature was correlated with sole temperature both before (r=0.93) and after (r=0.92) endovascular therapy. Before endovascular therapy, instep and sole temperatures were mean 30.8 °C (SD 3.0 °C) and mean 30.4 °C (SD 2.8 °C), respectively, and the foot temperature before endovascular therapy was mean 30.6 °C (SD 2.8 °C). After endovascular therapy, instep and sole temperatures were mean 32.6 °C (SD 2.9 °C) and mean 32.0 °C (SD 2.8 °C), respectively, and the foot temperature after endovascular therapy was mean 32.1 °C (SD 2.8 °C). The differences in instep (P=.005), sole (P=.009), and foot (P=.01) temperatures before and after endovascular therapy were statistically significant.

Prognosis of Wound Healing Time With Foot Temperature Change After Endovascular Therapy

Patients were categorized into those whose foot temperature increased by ≥2 °C (n=75) and those whose foot temperature increased by <2 °C (n=88) after endovascular therapy, in whom wound healing was achieved in 61 and 37 patients, respectively, over the course of the follow-up period. Wound healing occurred in a significantly lower percentage of patients whose foot

temperature increased by <2 °C (HR 0.41; 95% CI 0.27-0.63; *P*<.001) than in those whose foot temperature increased by ≥2 °C (Figure 3).

Figure 3. Wound healing time Kaplan-Meier curves (log-rank test *P*=.01; for trend *P*<.007). HR: hazard ratio.



Predictors of Wound Healing Time After Endovascular Therapy

Univariate analysis revealed that a nonambulatory status (HR 1.68, 95% CI 1.13-1.92, *P*=.04) and foot temperature increase >2 °C (HR 1.78, 95% CI 1.24-2.76, *P*=.02) were significantly predictive of wound healing within 180 days. Dialysis dependence was significantly negatively associated with wound healing within 180 days (HR 0.87, 95% CI 0.61-0.96, *P*=.03). In the multivariate Cox analysis, these associations remained significant (ambulatory status: HR 1.71, 95% CI, 1.18-1.91, *P*=.03; foot temperature increase ≥2 °C: HR 1.69; 95% CI, 1.21-2.67, *P*=.03; dialysis dependence: HR 0.81, 95% CI 0.63-0.92, *P*=.02).

Table 2. Univariate and multivariate analyses: predictors of wound healing at 180 days after endovascular therapy.

Variable	Univariate analysis		Multivariate analysis	
	HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
Age <80 years	1.54 (0.91-1.78)	.07	1.41 (0.82-1.86)	.15
Female	1.26 (0.74-1.56)	.45	1.15 (0.71-1.56)	.59
BMI[b]>18.5 kg/m^2	1.65 (0.85-1.88)	.32	1.46 (0.81-1.96)	.43
Not ambulatory	1.68 (1.13-1.92)	.04	1.71 (1.18-1.91)	.03
Diabetes	0.76 (0.45-1.46)	.22	0.87 (0.65-1.78)	.35
Hypertension	1.13 (0.61-1.52)	.56	1.19 (0.68-1.46)	.51
Dyslipidemia	1.27 (0.54-2.67)	.66	1.25 (0.61-2.28)	.48
Dialysis	0.87 (0.61-0.96)	.03	0.81 (0.63-0.92)	.02
Δ Foot temperature ≥2 °C	1.78 (1.24-2.76)	.02	1.69 (1.21-2.67)	.03

[a]HR: hazard ratio.

[b]BMI: body mass index.

Discussion

Principal Findings

We investigated mHealth-based monitoring of foot temperature change after endovascular therapy in patients with chronic foot ulcer. We demonstrated that a temperature increase >2 °C after endovascular therapy was strongly associated with faster healing time.

The endpoints currently used for endovascular therapy in clinical practice for critical limb ischemia vary. Wound bed temperature is crucial to wound healing—collagen deposition and the activities of late-phase inflammatory cells and fibroblasts are impeded when the temperature of the wound bed is lower than the core body temperature, leading to delayed healing [17]. It has been demonstrated in vitro that neutrophil, fibroblast, and epithelial cell activities decrease below a critical temperature of 33 °C [18]. Although the mean foot temperature after endovascular therapy in our cohort was less than this cut-off, our results showed that an increase in foot temperature is also an effective and reliable endpoint and may be a good surrogate endpoint for revascularization in critical limb ischemia.

However, increased local temperature is also a classic sign of wound infection. We, therefore, excluded patients with signs of local wound infection in the foot. Previous studies have shown that even mildly elevated temperatures may suggest occult infection that may lead to amputation [19]. Fierheller et al [20] found that integrating quantitative skin temperature measurements into routine wound assessment is a prompt and dependable means of quantifying the heat associated with skin infection and monitoring wound status. At-home self-monitoring with daily foot thermometry may thus be an effective adjunctive tool for preventing complications in individuals at high risk of lower-extremity ulceration and amputation. Combined with telehealth care, mHealth-based thermometry is a promising technology for treating chronic foot ulcer during the acute stage after endovascular therapy and the chronic stage of wound infection control.

mHealth is a popular trend, at present, that facilitates precise and individualized health care for various diseases. With the help of mobile devices, patients can easily record and transmit health-related parameters to a large database or to clinicians for consultation. Boodoo et al [21] demonstrated that patients were interested in mHealth for preventing and monitoring diabetic foot ulcers, even though some participants were not frequent users of mobile technology. However, not all mHealth-based monitoring options are appropriate for chronic foot ulcer treatment. Wearable sensors embedded in dressings or bandaging, for real-time monitoring of variables such as pH and temperature in diabetic foot ulcers, and telehealth platforms with medical-grade cameras are available on the market; however, they are often cumbersome and expensive. van Netten et al [22] suggested that images capture using mobile phones should not be used alone because they demonstrate low validity and reliability for the remote assessment of diabetic foot ulcers. Numerous mHealth studies of diabetes-related complications are now in progress, including studies on diabetes retinopathy [23] and chronic foot ulcer [21], which are associated with worsened cardiovascular prognoses. mHealth-based treatment continues to show potential in improving patient outcomes.

Adherence to frequent outpatient clinic visits may be difficult for patients with chronic foot ulcer who are older adults, bedridden or wheelchair-bound, and dependent on others for performing activities of daily living. Without intensive medical treatment, chronic foot ulcers are not likely to heal, which may result in further infection or amputation. The use of mHealth-based thermometry with telehealth can provide personalized evaluation and treatment and easy access to medical services: Lazo-Porras et al [24] reported that the implementation of foot thermometry through SMS prevented diabetic foot ulcers [24]. Recent reviews have found that SMS reminders were useful in several clinical scenarios, such as in enhancing adherence to antiretroviral [25], tuberculosis medications [26], and smoking cessation [27]. The system used in our study, a mobile phone–based thermometer with Bluetooth and app to transmit data automatically to the cloud, and would be easier for clinicians to use than SMS. The idea of combining mHealth and telehealth in caring for vulnerable patients is particularly attractive, and further clinical trials are needed to assess its efficacy.

Patients with diabetes and chronic foot ulcers exhibit poor prognoses in terms of cardiovascular outcomes; telehealth care can be used to provide comprehensive medical treatment beyond wound treatment alone. A previous study [28] has shown that telehealth care exerts positive effects in terms of disease control, including for hypertension, dyslipidemia, and diabetes, all of which are important predisposing factors for chronic foot ulcer, but chronic foot ulcer is strongly associated with concomitant cardiovascular disease, and therefore, its presence poses an increased risk of cardiovascular events [29]. In a previous study [30], we found that patients with cardiovascular disease who were receiving telehealth care had improved clinical outcomes.

Limitations

This study has several limitations. First, it was a single center cohort study with a limited number of patients; however, to the best of our knowledge, the study cohort is the largest cohort to date addressing the impact of an mHealth intervention on patients receiving endovascular therapy. Second, the existence of numerous confounding factors in our cohort study may have influenced the results. We used a multivariate Cox analysis model to diminish the possible confounding effects of other clinical factors, but conducting randomized trials would be advisable to verify our findings. Third, we measured the temperature before and after patients underwent endovascular therapy, with the assistance of study nurses; however, it would be challenging for patients in the target population (wheelchair-bound, bedridden) to use this modern device for daily at-home temperature self-monitoring. Development of a user-friendly interface, ideally with single-button activation, is necessary in the next generation mHealth-based device to meet the needs of these patients. Lastly, wound healing was defined by complete epithelialization of the wound as observed by the eye. Different wound locations and wound sizes on the foot may have influenced the rate of wound healing.

Conclusion

We demonstrated that increases in foot temperature by >2 °C after endovascular therapy, recorded with mHealth-based thermometers, were associated with better wound healing in patients with chronic foot ulcer. In the future, the mHealth-based thermometers may be combined with telehealth medicine for comprehensive care of diabetes patients with chronic foot ulcer. Multicenter studies, with longer follow-up, are needed to validate these results and provide guidelines for clinical practice.

Conflicts of Interest

None declared.

References

1. Rubio JA, Jiménez S, Álvarez J. Clinical characteristics and mortality in patients treated in a multidisciplinary diabetic foot unit. Endocrinología, Diabetes y Nutrición (English ed.) 2017 May;64(5):241-249. [doi: 10.1016/j.endien.2017.02.013]
2. Tuttolomondo A, Maida C, Pinto A. Diabetic foot syndrome: immune-inflammatory features as possible cardiovascular markers in diabetes. World J Orthop 2015 Jan 18;6(1):62-76 [FREE Full text] [doi: 10.5312/wjo.v6.i1.62] [Medline: 25621212]
3. Tuttolomondo A, Maida C, Pinto A. Diabetic foot syndrome as a possible cardiovascular marker in diabetic patients. J Diabetes Res 2015;2015:268390-268312 [FREE Full text] [doi: 10.1155/2015/268390] [Medline: 25883983]
4. Shishehbor MH, White CJ, Gray BH, Menard MT, Lookstein R, Rosenfield K, et al. Critical limb ischemia: an expert statement. J Am Coll Cardiol 2016 Nov 01;68(18):2002-2015 [FREE Full text] [doi: 10.1016/j.jacc.2016.04.071] [Medline: 27692726]
5. Aboyans V, Ricco JB, Bartelink MLEL, Björck M, Brodmann M, Cohnert T, et al. 2017 ESC guidelines on the diagnosis and treatment of peripheral arterial diseases, in collaboration with the European Society for Vascular Surgery. Russ J Cardiol 2018 Sep 09(8):164-221. [doi: 10.15829/1560-4071-2018-8-164-221]
6. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2017 Mar 21;135(12):e686-e725 [FREE Full text] [doi: 10.1161/CIR.0000000000000470] [Medline: 27840332]
7. Iida O, Soga Y, Hirano K, Kawasaki D, Suzuki K, Miyashita Y, et al. Long-term results of direct and indirect endovascular revascularization based on the angiosome concept in patients with critical limb ischemia presenting with isolated below-the-knee lesions. J Vasc Surg 2012 Feb;55(2):363-370.e5 [FREE Full text] [doi: 10.1016/j.jvs.2011.08.014] [Medline: 22051875]
8. Kawarada O, Yokoi Y, Higashimori A, Waratani N, Fujihara M, Kume T, et al. Assessment of macro- and microcirculation in contemporary critical limb ischemia. Catheter Cardiovasc Interv 2011 Dec 01;78(7):1051-1058. [doi: 10.1002/ccd.23086] [Medline: 21805568]
9. Kobayashi N, Hirano K, Nakano M, Muramatsu T, Tsukahara R, Ito Y, et al. Development and validation of a new scoring system to predict wound healing after endovascular therapy in critical limb ischemia with tissue loss. J Endovasc Ther 2015 Feb;22(1):48-56. [doi: 10.1177/1526602814564370] [Medline: 25775680]
10. Fukunaga M, Fujii K, Kawasaki D, Nishimura M, Horimatsu T, Saita T, et al. Vascular flow reserve immediately after infrapopliteal intervention as a predictor of wound healing in patients with foot tissue loss. Circ Cardiovasc Interv 2015 Jun;8(6):e002412. [doi: 10.1161/CIRCINTERVENTIONS.115.002412] [Medline: 26019143]
11. Utsunomiya M, Nakamura M, Nagashima Y, Sugi K. Predictive value of skin perfusion pressure after endovascular therapy for wound healing in critical limb ischemia. J Endovasc Ther 2014 Oct;21(5):662-670. [doi: 10.1583/14-4675MR.1] [Medline: 25290794]
12. Azuma N, Koya A, Uchida D, Saito Y, Uchida H. Ulcer healing after peripheral intervention-can we predict it before revascularization? Circ J 2014;78(8):1791-1800 [FREE Full text] [doi: 10.1253/circj.cj-14-0500] [Medline: 24998193]
13. Antonopoulos CN, Lazaris A, Venermo M, Geroulakos G. Predictors of wound healing following revascularization for chronic limb-threatening ischemia. Vasc Endovascular Surg 2019 Nov 12;53(8):649-657. [doi: 10.1177/1538574419868863] [Medline: 31405350]
14. Boezeman R, Becx B, van den Heuvel D, Ünlü Ç, Vos J, de Vries J. Monitoring of foot oxygenation with near-infrared spectroscopy in patients with critical limb ischemia undergoing percutaneous transluminal angioplasty: a pilot study. Eur J Vasc Endovasc Surg 2016 Nov;52(5):650-656 [FREE Full text] [doi: 10.1016/j.ejvs.2016.07.020] [Medline: 27614555]
15. Yavuz M, Ersen A, Hartos J, Lavery LA, Wukich DK, Hirschman GB, et al. Temperature as a causative factor in diabetic foot ulcers: a call to revisit ulceration pathomechanics. J Am Podiatr Med Assoc 2019 Sep;109(5):345-350. [doi: 10.7547/17-131] [Medline: 30427732]
16. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

17. Dini V, Salvo P, Janowska A, Di Francesco F, Barbini A, Romanelli M. Correlation between wound temperature obtained with an infrared camera and clinical wound bed score in venous leg ulcers. Wounds 2015 Oct;27(10):274-278 [FREE Full text] [Medline: 26479211]
18. Xia Z, Sato A, Hughes MA, Cherry GW. Stimulation of fibroblast growth in vitro by intermittent radiant warming. Wound Repair Regen 2000 Dec 25;8(2):138-144. [doi: 10.1046/j.1524-475x.2000.00138.x] [Medline: 10810040]
19. Lavery LA, Petersen BJ, Linders DR, Bloom JD, Rothenberg GM, Armstrong DG. Unilateral remote temperature monitoring to predict future ulceration for the diabetic foot in remission. BMJ Open Diabetes Res Care 2019 Aug 06;7(1):e000696 [FREE Full text] [doi: 10.1136/bmjdrc-2019-000696] [Medline: 31423317]
20. Fierheller M, Sibbald RG. A clinical investigation into the relationship between increased periwound skin temperature and local wound infection in patients with chronic leg ulcers. Adv Skin Wound Care 2010 Aug;23(8):369-79; quiz 380. [doi: 10.1097/01.ASW.0000383197.28192.98] [Medline: 20631603]
21. Boodoo C, Perry JA, Hunter PJ, Duta DI, Newhook SCP, Leung G, et al. Views of patients on using mhealth to monitor and prevent diabetic foot ulcers: qualitative study. JMIR Diabetes 2017 Sep 15;2(2):e22 [FREE Full text] [doi: 10.2196/diabetes.8505] [Medline: 30291089]
22. van Netten JJ, Clark D, Lazzarini PA, Janda M, Reed LF. The validity and reliability of remote diabetic foot ulcer assessment using mobile phone images. Sci Rep 2017 Aug 25;7(1):9480 [FREE Full text] [doi: 10.1038/s41598-017-09828-4] [Medline: 28842686]
23. Umaefulam V, Premkumar K. Impact of mobile health in diabetic retinopathy awareness and eye care behavior among Indigenous women. Mhealth 2020;6:14 [FREE Full text] [doi: 10.21037/mhealth.2019.12.01] [Medline: 32270006]
24. Lazo-Porras M, Bernabe-Ortiz A, Sacksteder KA, Gilman RH, Malaga G, Armstrong DG, et al. Implementation of foot thermometry plus mHealth to prevent diabetic foot ulcers: study protocol for a randomized controlled trial. Trials 2016 Apr 19;17(1):206 [FREE Full text] [doi: 10.1186/s13063-016-1333-1] [Medline: 27094007]
25. Shah R, Watson J, Free C. A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection. BMC Public Health 2019 Jul 09;19(1):915 [FREE Full text] [doi: 10.1186/s12889-019-6899-6] [Medline: 31288772]
26. Lester R, Park JJ, Bolten LM, Enjetti A, Johnston JC, Schwartzman K, et al. Mobile phone short message service for adherence support and care of patients with tuberculosis infection: Evidence and opportunity. J Clin Tuberc Other Mycobact Dis 2019 Aug;16:100108 [FREE Full text] [doi: 10.1016/j.jctube.2019.100108] [Medline: 31720432]
27. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
28. Bingham JM, Black M, Anderson EJ, Li Y, Toselli N, Fox S, et al. Impact of telehealth interventions on medication adherence for patients with type 2 diabetes, hypertension, and/or dyslipidemia: a systematic review. Ann Pharmacother 2021 May;55(5):637-649. [doi: 10.1177/1060028020950726] [Medline: 32815400]
29. Iida O, Takahara M, Soga Y, Suzuki K, Hirano K, Kawasaki D, et al. Efficacy of intravascular ultrasound in femoropopliteal stenting for peripheral artery disease with TASC II class A to C lesions. J Endovasc Ther 2014 Aug;21(4):485-492. [doi: 10.1583/14-4721R.1] [Medline: 25101575]
30. Hung C, Yu J, Lin Y, Chen Y, Huang C, Lee J, et al. Mortality benefit of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease: a longitudinal study. J Med Internet Res 2016 May 13;18(5):e102 [FREE Full text] [doi: 10.2196/jmir.5718] [Medline: 27177497]

Abbreviations

HR: hazard ratio
mHealth: mobile health
SMS: short message service

Edited by G Eysenbach; submitted 13.12.20; peer-reviewed by R Renner, R Ferrer-Peña; comments to author 05.01.21; revised version received 14.03.21; accepted 03.04.21; published 07.05.21.

Please cite as:
Lin DSH, Lee JK
Mobile Health–Based Thermometer for Monitoring Wound Healing After Endovascular Therapy in Patients With Chronic Foot Ulcer: Prospective Cohort Study
JMIR Mhealth Uhealth 2021;9(5):e26468
URL: https://mhealth.jmir.org/2021/5/e26468
doi:10.2196/26468
PMID:33960955

Original Paper

Usability and Accuracy of a Smartwatch for the Assessment of Physical Activity in the Elderly Population: Observational Study

Matteo Martinato[1], MSc, PhD; Giulia Lorenzoni[1], MSc, PhD; Tommaso Zanchi[1], BSc; Alessia Bergamin[1], MSc; Alessia Buratin[1,2,3], MSc; Danila Azzolina[1,4], MSc, PhD; Dario Gregori[1], MA, PhD

[1]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac-Thoracic-Vascular Sciences and Public Health, University of Padova, Padova, Italy

[2]Department of Biology, University of Padova, Padova, Italy

[3]Department of Molecular Medicine, University of Padova, Padova, Italy

[4]Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy

Corresponding Author:
Dario Gregori, MA, PhD
Unit of Biostatistics, Epidemiology and Public Health
Department of Cardiac-Thoracic-Vascular Sciences and Public Health
University of Padova
Via Leonardo Loredan, 18
Padova,
Italy
Phone: 39 049 827 5384
Fax: 39 027 004 45089
Email: dario.gregori@unipd.it

Abstract

Background: Regular physical activity (PA) contributes to the primary and secondary prevention of several chronic diseases and reduces the risk of premature death. Physical inactivity is a modifiable risk factor for cardiovascular disease and a variety of chronic disorders such as diabetes, obesity, hypertension, bone and joint diseases (eg, osteoporosis and osteoarthritis), depression, and colon and breast cancer. Population aging and the related increase in chronic diseases have a major impact on the health care systems of most Western countries and will produce an even more significant effect in the future. Monitoring PA is a valuable method of determining whether people are performing enough PA so as to prevent chronic diseases or are showing early symptoms of those diseases.

Objective: The aim of this study was to estimate the accuracy of wearable devices in quantifying the PA of elderly people in a real-life setting.

Methods: Participants aged 70 to 90 years with the ability to walk safely without any walking aid for at least 300 meters, who had no walking disabilities or episodes of falling while walking in the last 12 months, were asked to walk 150 meters at their preferred pace wearing a vívoactive HR device (Garmin Ltd) and actual steps were monitored and tallied by a researcher using a hand-tally counter to assess the performance of the device at a natural speed. A Bland-Altman plot was used to analyze the difference between manually counted steps and wearable device–measured steps. The intraclass correlation coefficient (ICC) was computed (with a 95% confidence interval) between step measurements. The generalized linear mixed-model (GLMM) ICCs were estimated, providing a random effect term (random intercept) for the individual measurements (gold standard and device). Both adjusted and conditional ICCs were computed for the GLMM models considering separately the effect of age, sex, BMI, and obesity. Analyses were performed using R software (R Foundation for Statistical Computing) with the rms package.

Results: A total of 23 females and 26 males were enrolled in the study. The median age of the participants was 75 years. The Bland-Altman plot revealed that, excluding one observation, all differences across measurements were in the confidence bounds, demonstrating the substantial agreement between the step count measurements. The results were confirmed by an ICC equal to .98 (.96-.99), demonstrating excellent agreement between the two sets of measurements.

Conclusions: The level of accuracy of wearable devices in quantifying the PA of elderly people in a real-life setting that was found in this study supports the idea of considering wrist-wearable nonmedical devices (widely available in nonspecialized stores) as reliable tools. Both health care professionals and informal caregivers could monitor the level of PA of their patients.

(JMIR Mhealth Uhealth 2021;9(5):e20966) doi:10.2196/20966

KEYWORDS

wearable devices; elderly; physical activity; smartwatches

Introduction

There is remarkable evidence that regular physical activity (PA) contributes to the primary and secondary prevention of several chronic diseases and reduces the risk of premature death [1]. There is a linear relation between the amount of PA and health benefits, such that the most physically active people are at the lowest risk [1]. However, the greatest improvements in health status are recorded when people who are the least fit become physically active [1].

Physical inactivity is a modifiable risk factor for cardiovascular disease and a variety of chronic disorders such as diabetes, obesity, hypertension, bone and joint diseases (eg, osteoporosis and osteoarthritis), depression, and colon and breast cancer [2]. In this framework, primary prevention plays a key role in the management of these diseases. Moreover, health promotion programs should target people of all ages, since the risk of developing chronic diseases starts in childhood and increases with age [1].

Population aging, and the related increase in chronic diseases, has a major impact on the health care systems of most Western countries and will produce an even more significant effect in the future. It has been calculated that, in 2001, chronic diseases accounted for approximately 60% of the 56.5 million total reported deaths in the world and for approximately 46% of the global burden of disease [3]. Almost half of chronic disease-related deaths are attributable to cardiovascular diseases. Obesity and diabetes are also showing worrying trends, not only because they already affect a large part of the population but because they have started to appear earlier in people's lives. Chronic diseases can be prevented with a healthy diet, avoidance of tobacco products, and regular PA [4]. Moreover, chronic diseases lead to a limitation of mobility and PA of affected persons, with a slow, progressive, and sometimes unnoticed entry mechanism, culminating in a reduction of autonomy. Therefore, monitoring PA is a valuable parameter for determining whether people are performing enough PA to prevent chronic diseases or if they are showing early symptoms of those diseases [5].

In recent years, several methods of monitoring PA and sedentary behavior have been proposed. Self-reporting is a simple tool for assessing PA via the completion of questionnaires, interviews, and surveys [6]. Alternatively, PA can be monitored by diaries or logs, where information on all forms of activity are recorded day by day. Those tools require a detailed description of the performed activity, including its intensity and duration. This method could produce useful health-related data, but the approach requires considerable worktime to produce standardized data. The recording of relevant data should be relatively simple and cover several days to avoid any potential bias [7].

Videorecording, adopting static cameras, wearable cameras, or low-cost motion-sensing input systems such as Kinect (Microsoft Corp), is another example of an autonomous data collection method [8]. Although this approach has a definite role in the assessment of activity patterns with the advantage of direct observation, it is unlikely to be practicable for large groups of individuals, requiring a great amount of resources to analyze and quantify videorecordings.

Heart rate monitors are low-cost tools, and heart rate can be used as a good quality proxy for PA, but it is not a precise indicator of energy expenditure unless proper individual calibration is performed. Gold standard techniques for measuring energy expenditure are based on the double-labeled water method or indirect calorimetry measuring oxygen uptake, carbon dioxide production, and cardiopulmonary parameters, but these techniques, although accurate, require specialized training and are expensive and not suitable for large-scale studies.

Advances in technology are facilitating researchers to quantify PA, and accelerometry-based activity monitors may be more suitable methods. Accelerometers are small and easy-to-use devices that track movement in 1 to 3 dimensions (ie, anteroposterior, vertical, and mediolateral). Using these tools, people can measure the frequency, intensity, and duration of PA. They are comfortable to wear, relatively inexpensive, and accurate compared with research-grade PA devices [9]. Accelerometers are technically more advanced than pedometers, and being multiaxial, they can measure horizontal, lateral, and vertical movements. These devices can be used to measure steps, activity counts, energy expenditure, posture, walking, and different intensities of movement. In addition, the reliability and validity of accelerometer data are generally high [10].

Recently, more attention has been paid to wrist-worn accelerometers. They are convenient and comfortable to wear, and patient compliance improves significantly while participating in studies requiring prolonged measurements. At the same time, these systems provide a high level of accuracy. Furthermore, the integration of additional motion sensors can be considered to increase the overall performance. Such integrated sensors may include gyroscopes, magnetometers, barometers, GPS devices, and physiological sensors (eg, heart rate) devised to improve the assessment and detection of specific indicators.

In this sense, wearable motion detectors might be the most promising technology for enabling an automatic, continuous, and long-term assessment of subjects in free-living environments. In addition, the obtained PA parameters can be shared with health care providers and insurance platforms to better describe behavioral patterns and functional ability in high-risk subjects, thus providing important feedback regarding the overall health status of an individual and even the prediction of potential adverse health events.

Nevertheless, few data are available on the accuracy and reliability of such devices, which are often built for active sports or athletes, in an elderly population. Some authors have validated this device on healthy older adults in studies including 20 participants [11].

In previous studies, wrist-worn devices showed poorer agreement to reference devices, suggesting that researchers should consider that not all consumer-level activity monitors are equal in terms of accuracy, design, and function [12].

Current public health recommendations for adults aged 65 years and older in general good health (moderate or vigorous intensity activity) could be seen challenging for many older people, and the health benefits of light intensity activities have not been defined even if light activities, including walking, account for the biggest part of daily activity in older populations [13].

Moreover, results of previous studies have been obtained in experimental settings and/or comparing device performance with data provided by reference instruments and in younger populations; they should be confirmed in other settings and with other devices [14] in older subjects to be considered for clinical application in the elderly.

PA research performed with validated but commercially available smartwatches holds the potential to address this gap with a more comprehensive assessment of the benefit of the overall amount of time spent ambulatory daily, and thus help shape future interventions specifically designed for increasing daily PA in older adults [13].

The study aimed to assess the feasibility of extending the use of these devices for monitoring the health of the elderly population and estimating the accuracy of wearable devices in quantifying the PA of elderly people on a larger sample of 49 subjects in a setting similar to a real-life condition excluding only individuals with a diagnosis of atrial fibrillation and current anticoagulant treatment.

Methods

Study Design

In this study, a commercially available smartwatch has been compared with direct observation of step counts, a metric successfully used in interventions to improve clinical outcomes [15] in a real-life, noninterventional, controlled setting (a walk at own pace in a daily attended location).

Aiming to perform a study that could be considered a starting point to extend the use of these devices for monitoring the health of the elderly population, a more protected experimental setting was chosen; however, the study setting was not fully controlled but resembled a real-life context: the proposal for participation in the study and the execution were immediate and conducted in a local market, a setting frequented by the participants. Moreover, the subjects freely walked at their own pace.

The study was conducted in flat areas previously marked with fixed distances on the ground using a Mini Measuring Wheel odometer (Group Silverline Ltd) during daily life circumstances in different cities in northeastern Italy. The path was linear and previously measured.

Participants met the following criteria: signed the informed consent form, aged between 70 and 90 years, able to walk safely and without any walking aid for at least 300 meters, no history of episodes of falling while walking in the last 12 months, no current diagnosis of atrial fibrillation, and no current anticoagulant treatment. Enrollment and the assessment were performed on the same day.

Participating subjects were asked to walk 150 meters at their preferred pace to assess the performance of the device at a natural speed [16]. Subject characteristics such as sex, age, weight, height, health conditions, and the number of steps on the path were collected.

Actual steps were monitored and tallied by a researcher using a hand-tally counter. The hand-tally count has been used as a criterion measure for manually measuring steps [17].

Study Device

In this study, a vívoactive HR smartwatch (Garmin Ltd) was programmed with participant sex, age, weight, and height and fitted on the left wrist according to the user manual. The device was designed for monitoring physical activity, especially outdoors, and has been validated in a real-life setting in other studies on the adult population [18]. The device tested in this study can also provide raw data regarding heart rate, number of steps, sleep quality, and activity (walking) session duration. In the literature, it has been demonstrated that the vívoactive HR was more accurate at reflecting step count across a broader range of walking cadences than other devices [19] also considering different age groups and during various walking conditions, even during slow walking [20].

The vívoactive HR is one of the few devices compliant with the new technological standards for physical activity monitors [21]. This device allows a detailed download of the raw data relating to each walking or training session, which allows accurate tracking of the physical activity of the elderly subject for health monitoring purposes. The activity monitor was started simultaneously with the start of the test. This time point was also recorded on the case report form by a different researcher to allow identification of the start point of the measurement [22]. After testing was completed, data were downloaded onto a personal computer via USB drive for postprocessing and analysis.

Sample Size

The sample size computation was performed using the method proposed by Bonett [23] for estimating intraclass correlation coefficients (ICCs). Two different step counting methods (a gold standard manual evaluation and wearable device counting) were considered to tailor the study design [24]. An ICC value of .80 was used as the expected agreement for the sample size computation, as indicated in the literature for other studies evaluating the agreement between device step counting and the manual counting gold standard [25]. A sample size of 49 subjects was considered sufficient to estimate the expected ICC with a precision of 0.10 based on a 95% confidence interval.

Statistical Analysis

Descriptive statistics were reported as medians and interquartile ranges for continuous variables and counts and percentages for categorical variables. The Wilcoxon Kruskal-Wallis test was performed for continuous variables, and the Pearson chi-square test was performed for categorical variables.

A Bland-Altman plot was used to analyze the difference between manually counted steps and wearable device–measured steps. The ICC was also computed (with a 95% confidence interval) between the step measurements.

Generalized linear mixed-model (GLMM) ICCs were also estimated, providing a random effect term (random intercept) for the individual measurements (gold standard and device). Both adjusted and conditional ICCs [26] were computed for the GLMM models considering separately the effects of age, sex, BMI, obesity (BMI >30), and session duration in minutes.

The adjusted ICC only considers the random effects in the computation, while the conditional ICC also takes the fixed effects variances into account and evaluates how much the covariate variable explains the portion of the variability in the grouping structure (random intercept).

The likelihood ratio test (LRT) was performed comparing the goodness of fit of the separate covariate-adjusted models with that of the intercept model (null model).

The covariates indicating significant goodness-of-fit improvement (ie, a significant covariate effect on the agreement between measures) in comparison with the null model were selected to perform a generalized linear model (GLM) on the number of misclassified steps (ie, the absolute value of the difference between the gold standard and device measurement). The negative binomial parametrization was considered to adjust the model estimates for overdispersion. The dispersion test was also performed as indicated in the literature [27]. The model fit was evaluated by reporting the residual Q-Q plots. Analyses were performed using R (R Foundation for Statistical Computing) [28] with the rms package [29].

Results

A total of 23 females and 26 males were enrolled in the study. The median age of the participants was 75 years (Table 1). A considerable proportion of participants included in the sample were obese (33/49, 67%), and the majority (20/33, 61%) were male. The step counts for both measurement criteria were greater for female participants than for male. Many older people could not participate due to walking problems or because they used walking aids. Many people approached said they did not have time to listen to information on the study, and this affected recruitment. The relatively younger subjects were found to be more willing to receive information on the study than older people. On the other hand, study participation was very high: in 50 people selected after checking the inclusion and exclusion criteria, only one refused to participate in the study.

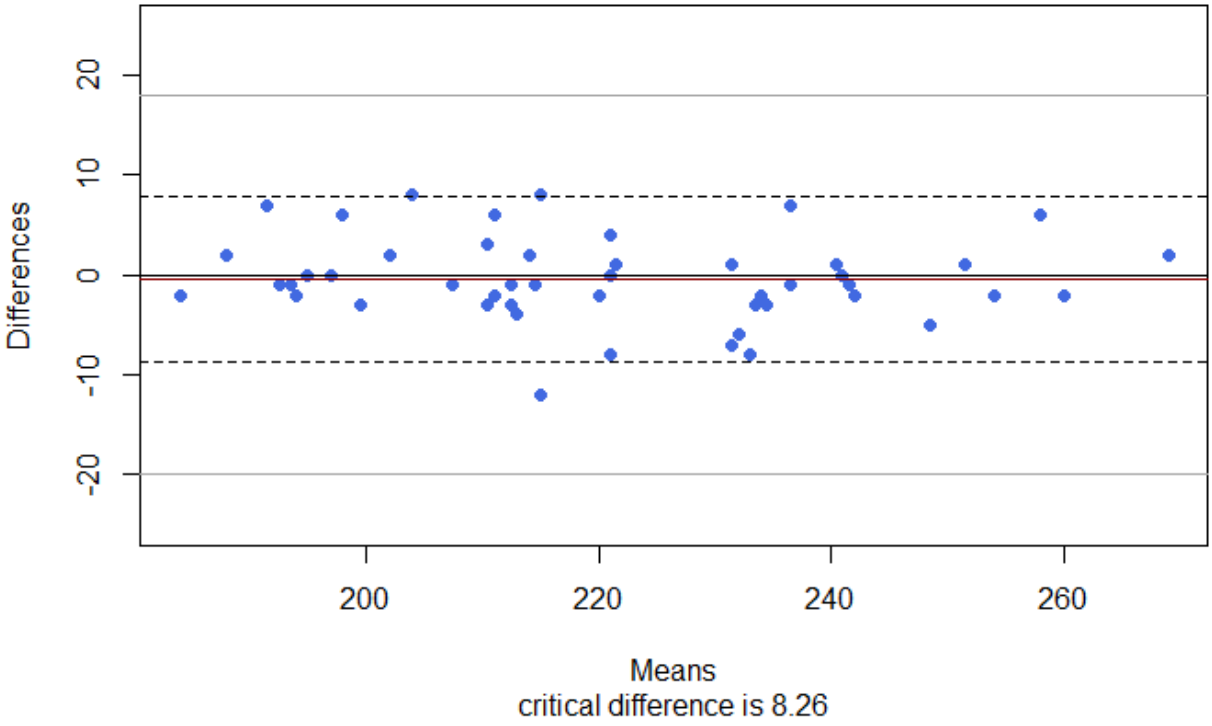
Table 1. Descriptive table of patient characteristics and step counts according to sex[a].

Characteristic	Females (n=23)	Males (n=26)	Combined (n=49)	*P* value
Patient demographic data				
Age in years, median (IQR[b])	75.00 (73.00-79.00)	72.00 (71.00-76.75)	74.00 (71.00-77.00)	.16
Weight (kg), median (IQR)	68.00 (62.50-76.50)	81.00 (75.00-90.00)	75.00 (68.00-85.00)	<.001
Height (m), median (IQR)	1.62 (1.58-1.66)	1.72 (1.70-1.77)	1.69 (1.62-1.73)	<.001
BMI, median (IQR)	25.00 (23.10-28.99)	27.21 (25.51-29.40)	26.99 (24.22-29.36)	.20
Obese (BMI >30), n (%)	11 (48)	22 (85)	33 (67)	.006
Step counts				
Gold standard, median (IQR)	232.0 (214.5-243.5)	211.0 (197.2-221.8)	217.0 (208.0-233.0)	.001
Device, median (IQR)	233.0 (216.0-247.0)	212.5 (195.5-221.0)	219.0 (208.0-236.0)	.002
Session duration (min), median (IQR)	2.000 (1.900-2.225)	1.950 (1.850-2.062)	1.967 (1.867-2.167)	.16

[a]The Wilcoxon Kruskal-Wallis test was performed for continuous variables, and the Pearson chi-square test was performed for categorical variables.
[b]IQR: interquartile range.

The Bland-Altman plot (Figure 1) reveals that, except for one observation, all the differences across measurements were in the confidence bounds, demonstrating substantial agreement between the step count measurements. Other research has been performed to validate wearable devices (Fitbit) on elderly subjects using the manual step count as the gold standard. Following this evidence, a clinically relevant lower limit of the agreement is –20 steps and an upper limit is 18 steps [30]. These limits have been represented in the Bland-Altman plot. All points lie within these boundaries indicating also that the range between the limits of agreement is narrow enough to represent nonclinically significant variation in the outcome. Results were also confirmed by an ICC equal to .98 (.96-.99), demonstrating excellent agreement between the two sets of measurements [25].

Figure 1. Bland-Altman plot of the difference between the gold standard daily steps and the wearable device measured daily steps (intraclass correlation coefficient .98, 95% CI 0.96-0.99). The red line indicates the mean difference (–0.4). The dotted lines indicate the Bland limit of agreement 1.96=*SD. The dark grey lines indicate a reasonable limit of agreement (-20; 18) as indicated in a study conducted to validate a Fitbit wearable device on an elderly population [31].



The GLMM conditional agreement analysis revealed that the ICC values, after controlling for the predictors, were different after adjusting for age and sex (Table 2). The LRT test revealed that age and sex were the covariates indicating significant improvement in model fit in comparison with the null model.

Table 2. Generalized linear mixed-model intraclass correlation coefficients (ICCs) have been estimated, providing a random effect term (random intercept) for individual measurements (gold standard and device). The adjusted ICC considers only the random effects, while the conditional ICC also takes the fixed-effect variances into account when evaluating how much the covariate variable explains the portion of the variation in the grouping structure (random intercept). The *P* value and chi-square test statistics are reported for the likelihood ratio test comparing the goodness of fit of the separate covariate-adjusted models with the intercept-only model (null model).

Model	Adjusted ICC[a]	Conditional ICC	Chi-square LRT[b]	*P* value (LRT)
Null model	.98	—[c]	—	—
Age in years	.98	.77	12	<.001
Sex	.98	.77	11.77	<.001
BMI	.98	.91	3.55	.06
Obesity	.98	.94	1.88	.17
Walking session duration	.98	.98	0.002	.90

[a]ICC: intraclass correlation coefficient.

[b]LRT: likelihood ratio test.

[c]Not applicable.

Considering the aforementioned results, sex and age were included as covariates to model the number of misclassified steps. A GLM with a negative binomial parametrization was considered for the estimation. This approach led to the adjustment of the estimates for the overdispersion component (φ). The dispersion test revealed that φ was equal to 2.11 and was significantly (P<.001) greater than 1 (φ = 1 indicates the absence of overdispersion in the data).

The multivariable negative binomial estimate reveals a nonsignificant age and sex effect on the number of misclassified steps (Figure 2). The residuals Q-Q plot indicates a good model fit performance for the negative binomial parametrizations (Figure 3).

Figure 2. Negative binomial model on the number of wearable device misclassified steps in comparison with the gold standard. The model predictions for the number of misclassified steps, according to sex and age, are reported in the plots. The model estimates, misclassification rate ratio with standard errors SE, and *P* values are reported in the tables. A .94 misclassification rate ratio indicates a 6% reduction in the misclassification rate (over the total number of steps) for a 1-year age increase. The overdispersion estimate is 2.12 (*P*<.001).

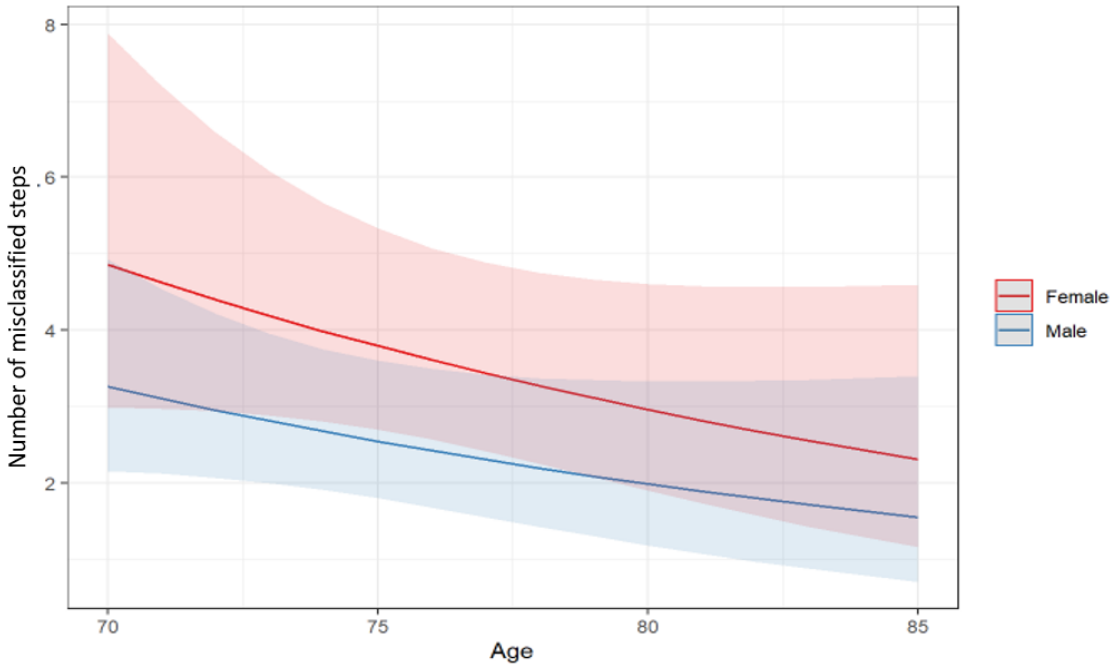
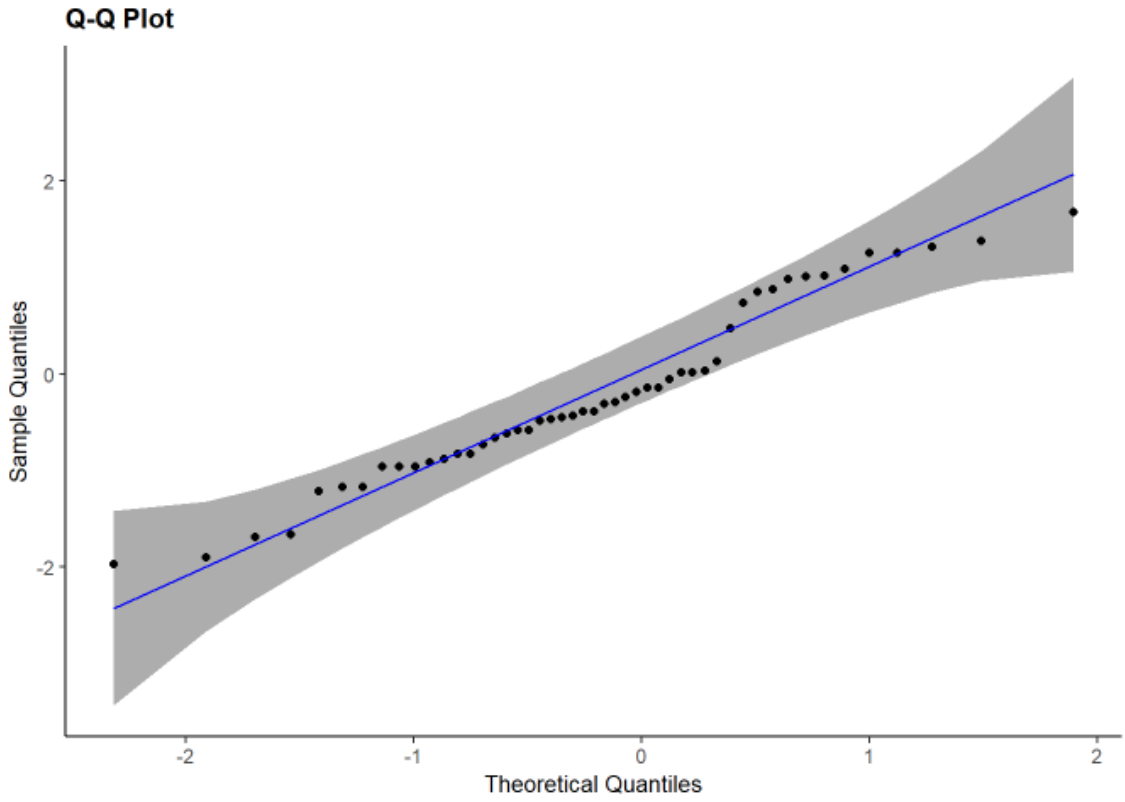


Figure 3. Residual Q-Q plots for the negative binomial models with 95% confidence bounds represented.



Discussion

Principal Findings

The purpose of the study was to determine the level of accuracy of a wearable nonmedical device that is widely available in nonspecialized stores in measuring PA in elderly individuals in a real-life setting. Collected data show that a wrist-wearable device is a reliable tool for measuring the level of PA in an elderly population in the context of daily life.

Given its potential and level of accuracy, a wearable device could be used by health care professionals to monitor the levels of PA in their patients. The application of these devices could be easily adopted in situations where there is a need to perform PA for the maintenance or improvement of patient health—for example, following orthopedic or major surgery or to achieve a beneficial increase in metabolism such as in the presence of diabetes or obesity [31-33]. Since these devices are commercially available, inexpensive, specific, and reliable, they could be used in today's health care environment, where the use of technological tools and adoption of telemedicine methods is becoming increasingly widespread. Wearable devices in health care are seen by older adults as a possible way to improve their health [5]. Indeed, they wish that devices were available in pharmacies, that they could learn about the devices from a health care professional, similar to other health monitoring systems (eg, blood pressure and blood glucose meters), and that health care professionals would use device-collected data.

Activity trackers are not taxed if prescribed in Canada [34], and in 2016 [5], a medical-grade exercise prescription device was recognized as a class 1 medical device. In the United States, a fitness tracker device is eligible for reimbursement when used to treat a medical condition such as obesity [35].

Are future challenges for wearable devices related to reimbursement by health care systems? Current research pipelines aim to collect data to validate their medical relevance to benchmark them against existing clinical solutions, reducing accuracy and reliability issues. Medically relevant clinical data should promote the integration of wearable devices into medical technologies, allowing a rethinking of cost covering by insurance companies and health care systems for clearly defined patient categories. Some examples of wearable systems are starting to be considered eligible for reimbursement [5]. The technology is advancing rapidly, and the market for wearable technology will expand significantly. Despite potential restraints and barriers, such data could cause a dramatic shift in the future of the life and health insurance industry. The evolution of wearable technology in health care is expected to revolutionize the health insurance industry, according to a new report from Timetric's Insurance Intelligence Center [36].

From a clinical perspective, it may be an important tool for studying the complete 24-hour activity cycle. A wearable device such as the one used in this study could also be easily adopted to measure 24-hour activity in elderly subjects for self-monitoring of spontaneous PA and/or sedentary behavior to prevent weight gain/regain in older adults and for self-monitoring of the effects of PA on self-efficacy and behavior in people with type 2 diabetes. Wearable devices could also allow closer PA monitoring of elderly individuals by both formal and informal caregivers, providing high efficiency for reacting to changes in behaviors.

Limitations

Study participants were aged 70 to 90 years, with median age being 75 years, a relatively young population considering actual life expectation. Data from this study should be applied with caution in older people. Subjects were selected excluding those needing any walking aid or with any walking disabilities; these study data cannot provide information regarding accuracy in assessing PA in subjects needing those aids or with walking disabilities. Moreover, further research developments are needed for extensive validation in real-life conditions. However, this study could represent a starting point for extending the use of these devices in elderly people, not only for clinical reasons but also for the health monitoring of a population that would benefit greatly from practicing a constant PA.

Conclusions

Results provided by this study could be considered a good starting point to plan further research considering the collection of different variables and more extensive observations. The device used in this study has the potential to capture the low levels of PA commonly performed by older adults which can be difficult to capture, including activities of daily living, but vital for maintaining health, independence, and quality of life with aging.

Final Remarks

The level of accuracy of wearable devices in quantifying the PA of older people in a real-life setting that was found in this study supports the idea of considering wrist-wearable nonmedical devices widely available in nonspecialized stores as reliable tools.

Conflicts of Interest

None declared.

References

1. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Impact of physical inactivity as a risk factor for chronic conditions: Australian Burden of Disease Study. Canberra: Australian Institute of Health and Welfare; 2017 Nov 22. URL: https://www.aihw.gov.au/getmedia/df392a65-8cf3-4c09-a494-4498ede2c662/aihw-bod-16.pdf.aspx?inline=true [accessed 2021-03-31]

3. World health report: shaping the future. Geneva: World Health Organization; 2003. URL: https://www.who.int/whr/2003/en/whr03_en.pdf [accessed 2021-03-31]
4. Willett W, Koplan J, Nugent R, Dusenbury C, Puska P, Gaziano T. Prevention of chronic disease by means of diet and lifestyle changes. In: Jamison D, Breman J, Measham A, Alleyne G, editors. Disease Control Priorities in Developing Countries. 2nd edition. Washington: The International Bank for Reconstruction and Development/The World Bank; 2006.
5. Tedesco S, Barton J, O'Flynn B. A review of activity trackers for senior citizens: research perspectives, commercial landscape and the role of the insurance industry. Sensors (Basel) 2017 Jun 03;17(6):1277 [FREE Full text] [doi: 10.3390/s17061277] [Medline: 28587188]
6. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. Sports Med 2001 Mar;31(2):91-100. [doi: 10.2165/00007256-200131020-00002] [Medline: 11227981]
7. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jul;37(3):197-206 [FREE Full text] [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]
8. Cudd P, Witte L. Harnessing the Power of Technology to Improve Lives. Amsterdam: IOS Press; 2017.
9. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 May;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
10. Evenson KR, Goto MM, 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]
11. Ehrler F, Weber C, Lovis C. Influence of pedometer position on pedometer accuracy at various walking speeds: a comparative study. J Med Internet Res 2016 Oct 06;18(10):e268 [FREE Full text] [doi: 10.2196/jmir.5916] [Medline: 27713114]
12. Farina N, Lowry RG. The validity of consumer-level activity monitors in healthy older adults in free-living conditions. J Aging Phys Act 2018 Jan 01;26(1):128-135. [doi: 10.1123/japa.2016-0344] [Medline: 28595019]
13. Schrack JA, Cooper R, Koster A, Shiroma EJ, Murabito JM, Rejeski WJ, et al. Assessing daily physical activity in older adults: unraveling the complexity of monitors, measures, and methods. J Gerontol A Biol Sci Med Sci 2016 Aug;71(8):1039-1048. [doi: 10.1093/gerona/glw026] [Medline: 26957472]
14. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
15. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
16. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
17. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. J Med Eng Technol 2017 Aug;41(6):480-485. [doi: 10.1080/03091902.2017.1333166] [Medline: 28675070]
18. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e10338 [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]
19. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018 Dec;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
20. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
21. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
22. Hurd WJ, Morrow MM, Kaufman KR. Tri-axial accelerometer analysis techniques for evaluating functional use of the extremities. J Electromyogr Kinesiol 2013 Aug;23(4):924-929 [FREE Full text] [doi: 10.1016/j.jelekin.2013.03.010] [Medline: 23642841]
23. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med 2002 May 15;21(9):1331-1335. [doi: 10.1002/sim.1108] [Medline: 12111881]
24. Giannakidou DM, Kambas A, Ageloussis N, Fatouros I, Christoforidis C, Venetsanou F, et al. The validity of two Omron pedometers during treadmill walking is speed dependent. Eur J Appl Physiol 2012 Jan;112(1):49-57. [doi: 10.1007/s00421-011-1951-y] [Medline: 21479653]
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
26. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 2017 Sep;14(134):20170213 [FREE Full text] [doi: 10.1098/rsif.2017.0213] [Medline: 28904005]

27. Cameron A, Trivedi PK. Regression-based tests for overdispersion in the Poisson model. J Econometrics 1990 Dec;46(3):347-364. [doi: 10.1016/0304-4076(90)90014-k]
28. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing URL: https://www.R-project.org/ [accessed 2021-03-31]
29. Harrell F. rms: Regression Modeling Strategies. New York: Springer International Publishing; 2015.
30. Paul SS, Tiedemann A, Hassett LM, Ramsay E, Kirkham C, Chagpar S, et al. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport Exerc Med 2015;1(1):e000013 [FREE Full text] [doi: 10.1136/bmjsem-2015-000013] [Medline: 27900119]
31. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
32. Chiang C, Chen K, Liu K, Hsu SJ, Chan C. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. Sensors (Basel) 2017 Feb 22;17(2):418 [FREE Full text] [doi: 10.3390/s17020418] [Medline: 28241434]
33. Meijer HA, Graafland M, Goslings JC, Schijven MP. Systematic review on the effects of serious games and wearable technology used in rehabilitation of patients with traumatic bone and soft tissue injuries. Arch Phys Med Rehabil 2018 Sep;99(9):1890-1899. [doi: 10.1016/j.apmr.2017.10.018] [Medline: 29138050]
34. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
35. FSA Store. URL: https://fsastore.com/FSA-Eligibility-List/F/Fitness-Tracker-E792.aspx [accessed 2021-04-14]
36. Johnson D. How wearable devices could disrupt the insurance industry. Insurance J. 2015 May 06. URL: https://www.insurancejournal.com/news/national/2015/05/06/367014.htm [accessed 2021-04-14]

Abbreviations

GLMM: generalized linear mixed-model
GLM: generalized linear model
ICC: intraclass correlation coefficient
LRT: likelihood ratio test
PA: physical activity

Edited by L Buis; submitted 02.06.20; peer-reviewed by MJ Toledo; comments to author 14.07.20; revised version received 02.10.20; accepted 24.02.21; published 05.05.21.

Please cite as:
Martinato M, Lorenzoni G, Zanchi T, Bergamin A, Buratin A, Azzolina D, Gregori D
Usability and Accuracy of a Smartwatch for the Assessment of Physical Activity in the Elderly Population: Observational Study
JMIR Mhealth Uhealth 2021;9(5):e20966
URL: https://mhealth.jmir.org/2021/5/e20966
doi:10.2196/20966
PMID:33949953

Original Paper

Exploring Breaks in Sedentary Behavior of Older Adults Immediately After Receiving Personalized Haptic Feedback: Intervention Study

Sofie Compernolle[1,2], PhD; Delfien Van Dyck[1], PhD; Greet Cardon[1], PhD; Ruben Brondeel[1,2], PhD

[1]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[2]Research Foundation Flanders, Brussels, Belgium

Corresponding Author:
Sofie Compernolle, PhD
Department of Movement and Sports Sciences
Ghent University
Watersportlaan 2
Ghent, 9000
Belgium
Phone: 32 92646323
Email: sofie.compernolle@ugent.be

Abstract

Background: "Push" components of mobile health interventions may be promising to create conscious awareness of habitual sedentary behavior; however, the effect of these components on the near-time, proximal outcome, being breaks in sedentary behavior immediately after receiving a push notification, is still unknown, especially in older adults.

Objective: The aims of this study are to examine if older adults break their sedentary behavior immediately after receiving personalized haptic feedback on prolonged sedentary behavior and if the percentage of breaks differs depending on the time of the day when the feedback is provided.

Methods: A total of 26 Flemish older adults (mean age 64.4 years, SD 3.8) wore a triaxial accelerometer (Activator, PAL Technologies Ltd) for 3 weeks. The accelerometer generated personalized haptic feedback by means of vibrations each time a participant sat for 30 uninterrupted minutes. Accelerometer data on sedentary behavior were used to estimate the proximal outcome, which was sedentary behavior breaks immediately (within 1, 3, and 5 minutes) after receiving personalized haptic feedback. Generalized estimating equations were used to investigate whether or not participants broke up their sedentary behavior immediately after receiving haptic feedback. A time-related variable was added to the model to investigate if the sedentary behavior breaks differed depending on the time of day.

Results: A total of 2628 vibrations were provided to the participants during the 3-week intervention period. Of these 2628 vibrations, 379 (14.4%), 570 (21.7%), and 798 (30.4%) resulted in a sedentary behavior break within 1, 3 and 5 minutes, respectively. Although the 1-minute interval did not reveal significant differences in the percentage of breaks depending on the time at which the haptic feedback was provided, the 3- and 5-minute intervals did show significant differences in the percentage of breaks depending on the time at which the haptic feedback was provided. Concretely, the percentage of sedentary behavior breaks was significantly higher if personalized haptic feedback was provided between noon and 3 PM compared to if the feedback was provided between 6 and 9 AM (odds ratio 1.58, 95% CI 1.01-2.47, within 3 minutes; odds ratio 1.78, 95% CI 1.11-2.84, within 5 minutes).

Conclusions: The majority of haptic vibrations, especially those in the morning, did not result in a break in the sedentary behavior of older adults. As such, simply bringing habitual sedentary behavior into conscious awareness seems to be insufficient to target sedentary behavior. More research is needed to optimize push components in interventions aimed at the reduction of the sedentary behavior of older adults.

Trial Registration: ClinicalTrials.gov NCT04003324; https://clinicaltrials.gov/ct2/show/NCT04003324

(JMIR Mhealth Uhealth 2021;9(5):e26387) doi:10.2196/26387

KEYWORDS

tactile feedback; sitting behavior; sedentary behavior; older adults; mHealth intervention; self-monitoring

Introduction

Evidence shows that older adults (aged ≥60 years) are the most sedentary segment of the population [1]. They spend approximately 80% of their awake time (8-12 hours per day) in sedentary activities [2]. This finding is alarming, as prolonged sedentary behavior has been associated with increased risk for negative health outcomes, such as frailty, type 2 diabetes, and all-cause mortality [3]. To date, the number of interventions targeting sedentary behavior in older adults has been limited. Only 9 interventions could be detected in a systematic review [4], of which only 1 was a mobile health (mHealth) intervention [5]. This finding contrasts sharply with the range of mHealth interventions aimed at the promotion of physical activity [6], and it is also very disappointing, as a recent meta-analysis showed a significantly higher decrease in sedentary behavior following mHealth interventions compared to traditional interventions in all age groups [7]. The superiority of mHealth interventions over traditional interventions may be explained by the fact that a large amount of sedentary behavior is contextually triggered and automatic (ie, it involves little reasoning and is performed without conscious decision-making) [8]. In contrast, physical activity is regulated by controlled processes, such as intentions, values, and beliefs. Undesired automatic behavior can be disrupted by bringing the behavior and its context into conscious awareness—for example, by means of self-monitoring [9]. Self-monitoring can be easily integrated in mHealth interventions; therefore, these interventions offer great potential to reduce sedentary behavior.

mHealth interventions generally consist of multiple intervention components. Some are “pull” components, which require individuals to access the component on their mobile device at moments when they decide they need help. Others are “push” components, which are initiated by the intervention and are delivered via haptic vibrations, notifications, or text messages [10]. Previous efficacy studies mainly investigated whether the combination of pull and push components resulted in a reduction of total sedentary time at the end of the intervention. This reduction can be considered to be the desired distal outcome of the intervention. However, as push components may best facilitate the process of bringing habitual sedentary behavior and its context into conscious awareness [11,12], it may be more pertinent to investigate the near-time, proximal effect of these push components (ie, breaks in sedentary behavior immediately after receiving a push notification) compared to their effect on total sedentary time. To date, little effort has focused on examining these near-time, proximal effects [13]; only one study could be found in the literature [14]. In this study, 86 office workers received persuasive text messages to break up their sedentary behavior after 30 minutes of uninterrupted computer time. The results showed a steep decline in sedentary behavior in the 30 minutes following a text message compared to a control group [14]. To our knowledge, no studies are available on the proximal outcomes of push components aimed at the reduction of sedentary behavior in older adults.

Therefore, the aims of this exploratory study were (1) to examine if older adults break their sedentary behavior immediately after receiving personalized haptic feedback, and (2) to investigate if those breaks differed depending on the time of the day when the feedback was delivered. This investigation may be important, as push components that are offered at an inappropriate time may lead to burden and disengagement [15].

Methods

Study Design

This study reports on part of a larger mixed methods study being conducted to evaluate a self-monitoring mHealth intervention to reduce sedentary behavior in older adults [16]. The study was registered at ClinicalTrials.gov (identification number: NCT04003324) and was approved by the Committee of Medical Ethics of the Ghent University Hospital (Belgian registration number 2019/0398). All participants provided written informed consent.

Participants, Procedure, and Intervention

Participant recruitment was conducted in Flanders between February and March 2019 using convenience sampling (ie, Facebook advertisements and an existing database [17]). To be eligible for the current study, participants needed to (1) be at least 60 years old, (2) be Dutch-speaking, (3) be able to walk 100 meters without severe difficulties, and (4) have a smartphone. A detailed description of the study procedure has been published elsewhere [16]. Briefly, baseline data, including sociodemographic characteristics, were collected before the start of the intervention. Subsequently, the self-monitoring mHealth intervention was introduced to the participants. The intervention consisted of general sedentary behavior information and visual and haptic feedback on the participants’ sedentary behavior. General sedentary behavior information was provided to participants by means of a 10-minute presentation. The presentation was given by an expert in the field during the second home visit. Visual and haptic feedback were provided using a novel, validated triaxial accelerometer—the Activator (PAL Technologies Ltd) [18]. The Activator was worn during waking hours on the front of the thigh, either in a pants pocket or attached with an elastic band to clothing covering the upper thigh (eg, trousers, jeans, shorts, leggings, tights, or dresses) [19]. Real-time visual feedback and a 7-day historical overview were presented through a smartphone app via a Bluetooth connection. Haptic feedback was provided by a strong but comfortable vibration of the Activator accelerometer itself each time a participant sat for 30 uninterrupted minutes. Participants were instructed to break up their sedentary behavior each time they received a haptic vibration.

Measures

The participants’ sociodemographic characteristics were administered using a structured interview and included age, gender, family situation, educational level, weight, and height. Sedentary behavior after receiving personalized haptic feedback

was collected using the Activator device. The Activator device collected triaxial accelerometer data about thigh position and accelerations and processed the data via proprietary algorithms (Intelligent Activity Classification, PAL Technologies) to determine the wearer's body posture (ie, sitting/lying and upright) and stepping speed. Accelerometer data were stored on the cloud server of PAL Technologies and used to estimate the proximal outcome, which was a break in sedentary behavior immediately (within 1, 3, and 5 minutes) after receiving personalized haptic feedback. Time-related characteristics of when the haptic feedback was provided were extracted from the system usage data of the Activator and categorized into the following six categories: 6-9 AM, 9 AM-noon, noon-3 PM, 3-6 PM, 6-9 PM, 9 PM-midnight.

Data Cleaning and Statistical Analyses

Descriptive statistics of the sample and of the personalized haptic feedback were summarized as proportions, means, and standard deviations. Participants' sedentary behavior within 1, 3, and 5 minutes after receiving haptic feedback was extracted from the accelerometer data and dichotomized as whether or not participants had broken up their sedentary behavior. If a participant broke up their sedentary behavior in the first minute after receiving a notification, this was taken into account in all three time frames. As each participant received haptic feedback multiple times, observations were nested for the participants. Generalized estimating equations, including a random intercept, were applied to investigate whether participants broke up their sedentary behavior immediately after receiving haptic feedback. The time-related variable was added to the model as a fixed effect to investigate if the sedentary behavior breaks differed depending on the time of day. Nonstandardized regression coefficients (β) and 95% confidence intervals were reported as effect estimates. If sedentary behavior breaks were observed, proportions were requested to determine the duration of the breaks. Analyses were performed in SPSS, version 25 (IBM Corporation).

Results

Descriptive Statistics of the Participants

The participant characteristics are presented in Table 1. Half of the participants were female, and the average age was 64.4 years (SD 3.8). The majority of the participants were highly educated and were married or lived with a partner. The participants' mean BMI was 25.2 kg/m^2 (SD 3.8).

Table 1. Participant characteristics (N=26).

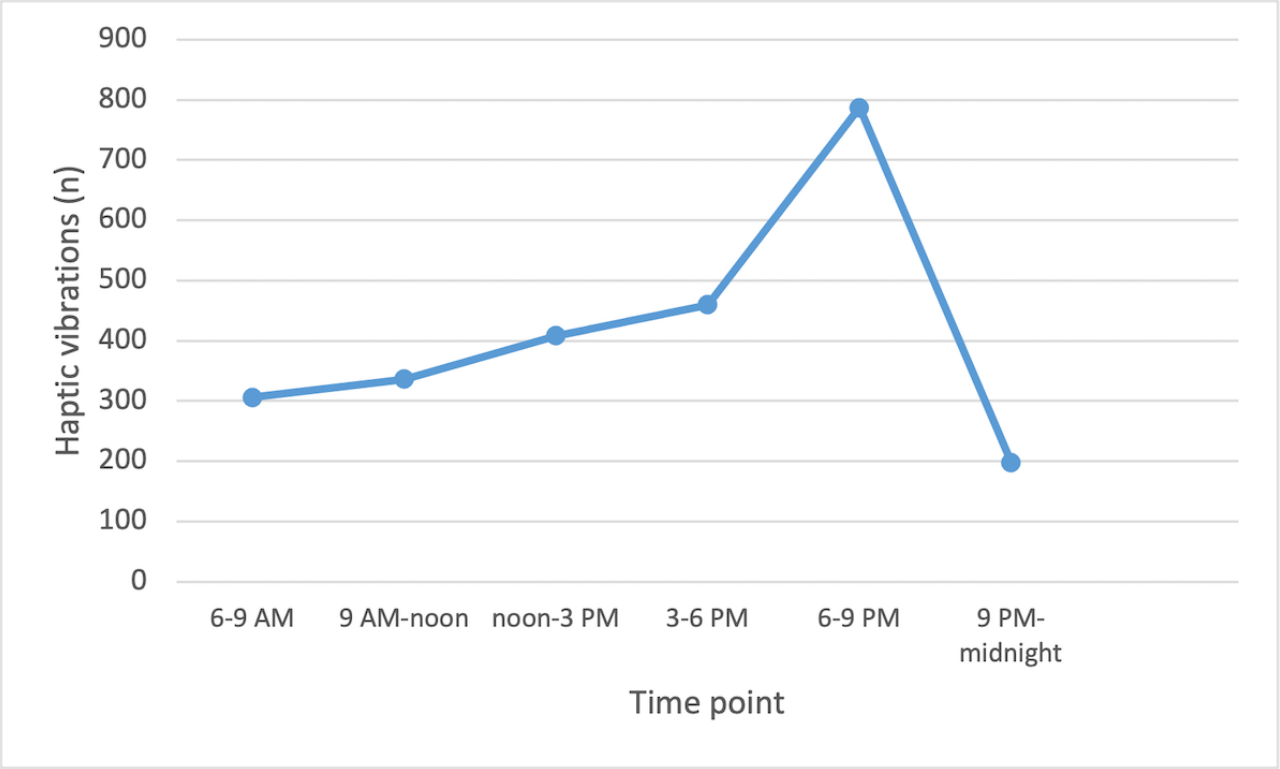
Sociodemographic characteristic	Value
Gender	
Men, n (%)	13 (50)
Women, n (%)	13 (50)
Age (years)	
Mean (SD), range	64.4 (3.8), 60.0-76.0
Young older adults (<65), n (%)	14 (54)
Older adults (≥65), n (%)	12 (46)
Educational level, n (%)	
Secondary education	11 (42)
College or university	15 (58)
Family situation, n (%)	
No partner (ie, single, widowed)	4 (15)
Partner but living separately	1 (4)
Married or living with a partner	21 (81)
BMI (kg/m^2)	
Mean (SD), range	25.2 (3.8), 19.7-32.3
Healthy weight (<25), n (%)	16 (62)
Overweight (25-29.9), n (%)	6 (23)
Obese (>30), n (%)	4 (15)

Descriptive Statistics of the Personalized Haptic Feedback

A total of 2628 vibrations were provided to the participants during the 3-week intervention period. The highest number of vibrations was provided between 6 and 9 PM, whereas the lowest number of vibrations was provided between 6 and 9 AM (306 and 787, respectively) (see Figure 1). Considerable differences were observed in the number of vibrations between participants, ranging from 3-258 vibrations during the 3-week intervention period. The median number of vibrations that participants received per day varied from 0-7. Detailed

information on the participants' personalized haptic feedback is provided in Multimedia Appendix 1.

Figure 1. Number of haptic vibrations at different time points.



Breaks in the Sedentary Behavior of Older Adults Immediately After Receiving Personalized Haptic Feedback

Of the 2628 personalized haptic vibrations, 379 (14.4%), 570 (21.7%), and 798 (30.4%) resulted in a sedentary behavior break within 1, 3, and 5 minutes, respectively. Although the 1-minute interval did not reveal significant differences in the percentage of breaks depending on the time at which the haptic feedback was provided, the 3 and 5-minute intervals did show significant differences in the percentage of breaks depending on the time at which the haptic feedback was provided. Concretely, the percentage of sedentary behavior breaks was significantly higher if personalized haptic feedback was provided between noon and 3 PM (3- and 5-minute intervals), compared to if the feedback was provided between 6 and 9 AM (see Table 2 and Figure 2). The duration of the breaks observed in the older adults' sedentary behavior is summarized in Table 3.

Table 2. Sedentary behavior breaks by time of the day when the vibrations were provided.

Time of day	1 minute		3 minutes		5 minutes	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
6-9 AM	1.00 (reference)	N/A[b]	1.00 (reference)	N/A	1.00 (reference)	N/A
9 AM-noon	1.38 (0.75-2.54)	.31	1.46 (0.81-2.62)	.21	1.42 (0.80-2.54)	.23
Noon-3 PM	1.26 (0.78-2.03)	.34	*1.58 (1.01-2.47)* [c]	*.05*	*1.78 (1.11-2.84)*	*.02*
3-6 PM	1.58 (0.85-2.94]	.15	1.62 (0.93-2.81)	.09	1.37 (0.80-2.36)	.25
6-9 PM	1.92 (0.74-2.25)	.36	1.41 (0.80-2.48)	.24	1.25 (0.71-2.21)	.44
9 PM-midnight	1.46 (0.62-3.44)	.39	1.66 (0.86-3.23)	.13	1.38 (0.67-2.87)	.38

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]Italic text indicates the most significant time periods.

Figure 2. Percentages of vibrations followed by a break in sedentary behavior by time of day.

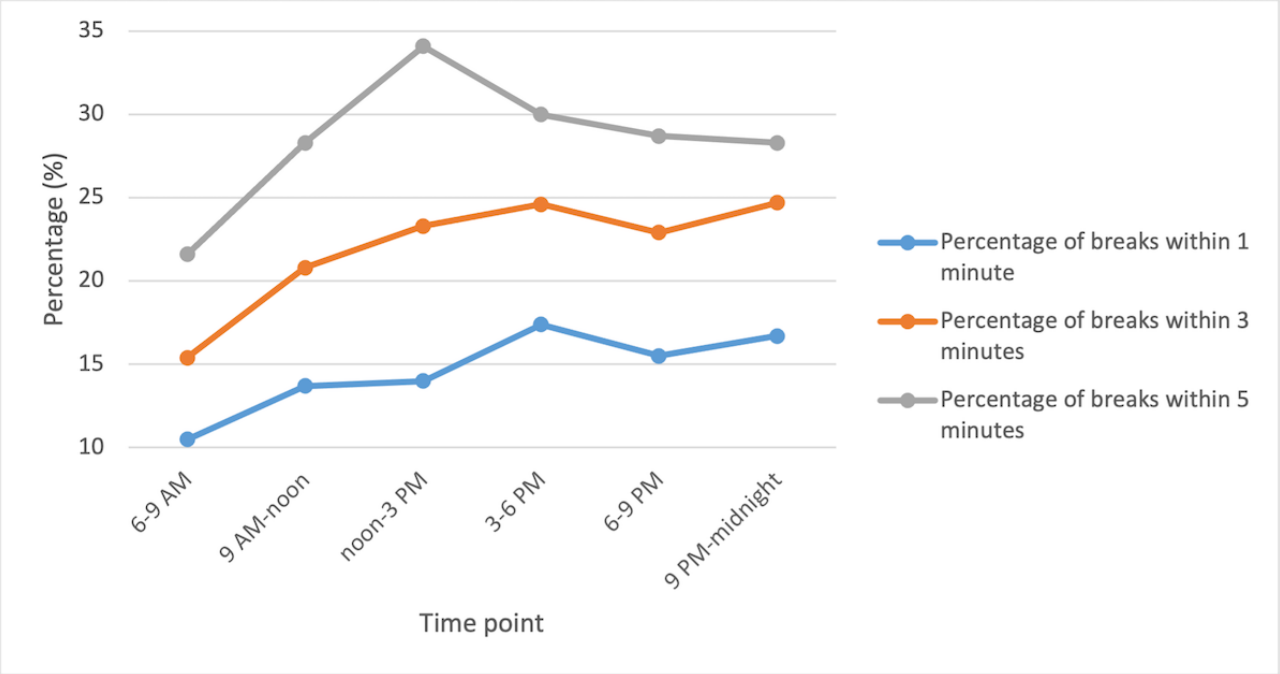


Table 3. Duration of the sedentary behavior breaks.

Duration of break (minutes)	Vibrations followed by a break within 3 minutes (n=546), n (%)	Vibrations followed by a break within 5 minutes (n=714), n (%)
<1	246 (45.1)	279 (38.7)
1-2	121 (22.1)	127 (17.6)
2-3	179 (32.8)	93 (12.9)
3-5	N/A[a]	215 (29.9)

[a]N/A: not applicable.

Discussion

This is the first study investigating breaks in the sedentary behavior of older adults immediately after receiving personalized haptic feedback. The results are rather disappointing, as less than 1 in 3 vibrations resulted in a sedentary behavior break within 5 minutes. Moreover, more than half of the breaks lasted less than 2 minutes, which may be too short to achieve health benefits [20]. Consequently, it can be concluded that simply increasing awareness of habitual sedentary behavior by means of personalized haptic feedback is insufficient to stimulate older adults to break their sedentary behavior. Although underlying reasons to ignore the personalized haptic feedback were not examined, previous research showed that many older adults lack the motivation (or capabilities) to break their sedentary behavior [21]. Therefore, more effort should be made to enhance older adults' motivation to break their sedentary behavior.

Our results also suggested that the percentage of breaks differed depending on the time of the day when the haptic feedback was provided (at least when analyzing the 3- and 5-minute intervals). Based on previous ecological momentary assessment studies, showing that older adults' fatigue increases throughout the day [22,23], it was expected that participants would be more likely to respond to the haptic feedback in the morning compared to in the afternoon and the evening. However, our results showed the opposite. Older adults were more likely to break their sedentary behavior in the afternoon compared to in the morning. It is possible that the moments when the participants are sedentary in the morning are "necessary moments of rest," as our results showed that they are much more active in the morning. On the other hand, in the afternoon, older adults are much more sedentary and probably thus more motivated to break their sedentary behavior after receiving a haptic vibration [1]. If this result can be replicated in future research, haptic vibrations, notifications, or text messages aimed at the reduction of older adults' sedentary behavior should preferably be provided in the afternoon to be successful.

The main strength of the current study is its innovativeness. As far as we know, no previous studies have investigated older adults' breaks in sedentary behavior immediately after receiving haptic feedback. The most relevant limitations are the small sample size and the simplicity of the study design. The small sample size hinders the investigation of individual differences (eg, sociodemographic and behavioral characteristics) between participants who often broke their sedentary behavior and those who did not. The simplicity of the study design prevents us from drawing firm conclusions on the causality of the

association between the push components and breaks in sedentary behavior. As a vibration was provided each time a participant was sedentary for 30 minutes, it remains unclear what the response would have been (ie, break or no break) if no vibration was given. Therefore, the use of micro-randomized trials is recommended to confirm and further elaborate the current findings.

Conclusion

The majority of haptic vibrations, especially those received in the morning, did not result in a break in older adults' sedentary behavior. As such, simply bringing habitual sedentary behavior into conscious awareness seems to be insufficient to target sedentary behavior. More research is needed to optimize push components in interventions aimed at the reduction of sedentary behavior in older adults.

Acknowledgments

RB (FWO18/PDO/029) and SC (FWO20/PDJ/088) are supported by a postdoctoral fellowship of the Research Foundation Flanders.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of vibrations per participant per day.
[DOCX File , 17 KB - mhealth_v9i5e26387_app1.docx]

References

1. Leask CF, Harvey JA, Skelton DA, Chastin SF. Exploring the context of sedentary behaviour in older adults (what, where, why, when and with whom). Eur Rev Aging Phys Act 2015 Oct 7;12(1):4-8 [FREE Full text] [doi: 10.1186/s11556-015-0146-7] [Medline: 26865868]
2. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 01;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
3. de Rezende LFM, Rodrigues Lopes M, Rey-López JP, Matsudo VKR, Luiz ODC. Sedentary behavior and health outcomes: an overview of systematic reviews. PLoS One 2014 Aug 21;9(8):e105620 [FREE Full text] [doi: 10.1371/journal.pone.0105620] [Medline: 25144686]
4. Chastin S, Gardiner P, Ashe M, Harvey J, Leask C, Balogun S, et al. Interventions for reducing sedentary behaviour in community-dwelling older adults. Cochrane Database Syst Rev 2017:9. [doi: 10.1002/14651858.cd012784]
5. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR mHealth uHealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
6. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019 Aug 01;27(4):565-593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
7. 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 18;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
8. Maher JP, Conroy DE. A dual-process model of older adults' sedentary behavior. Health Psychol 2016 Mar;35(3):262-272. [doi: 10.1037/hea0000300] [Medline: 26690644]
9. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Computers in Human Behavior 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
10. Luers B, Klasnja P, Murphy S. Standardized effect sizes for preventive mobile health interventions in micro-randomized trials. Prev Sci 2019 Jan 9;20(1):100-109 [FREE Full text] [doi: 10.1007/s11121-017-0862-5] [Medline: 29318443]
11. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63-16 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
12. Conroy DE, Maher JP, Elavsky S, Hyde AL, Doerksen SE. Sedentary behavior as a daily process regulated by habits and intentions. Health Psychol 2013 Nov;32(11):1149-1157 [FREE Full text] [doi: 10.1037/a0031629] [Medline: 23477579]
13. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S(Suppl):1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]

14. van Dantzig S, Geleijnse G, van Halteren AT. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246. [doi: 10.1007/s00779-012-0588-0]
15. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul 19;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]
16. Compernolle S, Cardon G, van der Ploeg HP, Van Nassau F, De Bourdeaudhuij I, Jelsma JJ, et al. Engagement, acceptability, usability, and preliminary efficacy of a self-monitoring mobile health intervention to reduce sedentary behavior in Belgian older adults: mixed methods study. JMIR mHealth uHealth 2020 Oct 29;8(10):e18653 [FREE Full text] [doi: 10.2196/18653] [Medline: 33118951]
17. Lakerveld J, Ben Rebah M, Mackenbach JD, Charreire H, Compernolle S, Glonti K, et al. Obesity-related behaviours and BMI in five urban regions across Europe: sampling design and results from the SPOTLIGHT cross-sectional survey. BMJ Open 2015 Oct 27;5(10):e008505 [FREE Full text] [doi: 10.1136/bmjopen-2015-008505] [Medline: 26507356]
18. Gill JMR, Hawari NSA, Maxwell DJ, Louden D, Mourselas N, Bunn C, et al. Validation of a Novel Device to Measure and Provide Feedback on Sedentary Behavior. Med Sci Sports Exerc 2018 Mar;50(3):525-532 [FREE Full text] [doi: 10.1249/MSS.0000000000001458] [Medline: 29040225]
19. Jelsma JGM, Renaud LR, Huysmans MA, Coffeng JK, Loyen A, van Nassau F, et al. The Dynamic Work study: study protocol of a cluster randomized controlled trial of an occupational health intervention aimed at reducing sitting time in office workers. BMC Public Health 2019 Feb 13;19(1):188 [FREE Full text] [doi: 10.1186/s12889-019-6467-0] [Medline: 30760231]
20. Kehler DS, Clara I, Hiebert B, Stammers AN, Hay J, Schultz A, et al. The association between patterns of physical activity and sedentary time with frailty in relation to cardiovascular disease. Aging Med (Milton) 2019 Mar 19;2(1):18-26 [FREE Full text] [doi: 10.1002/agm2.12059] [Medline: 31942509]
21. Compernolle S, De Cocker K, Cardon G, De Bourdeaudhuij I, Van Dyck D. Older adults' perceptions of sedentary behavior: a systematic review and thematic synthesis of qualitative studies. Gerontologist 2020 Nov 23;60(8):572-582. [doi: 10.1093/geront/gnz127] [Medline: 31670766]
22. Dunton GF, Atienza AA, Castro CM, King AC. Using ecological momentary assessment to examine antecedents and correlates of physical activity bouts in adults age 50+ years: a pilot study. Ann Behav Med 2009 Dec 6;38(3):249-255 [FREE Full text] [doi: 10.1007/s12160-009-9141-4] [Medline: 20052568]
23. Curran SL, Beacham AO, Andrykowski MA. Ecological momentary assessment of fatigue following breast cancer treatment. J Behav Med 2004 Oct;27(5):425-444. [doi: 10.1023/b:jobm.0000047608.03692.0c]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 09.12.20; peer-reviewed by R Krukowski, M Sobolev, B Nanda; comments to author 21.01.21; revised version received 24.02.21; accepted 15.04.21; published 10.05.21.

Please cite as:
Compernolle S, Van Dyck D, Cardon G, Brondeel R
Exploring Breaks in Sedentary Behavior of Older Adults Immediately After Receiving Personalized Haptic Feedback: Intervention Study
JMIR Mhealth Uhealth 2021;9(5):e26387
URL: https://mhealth.jmir.org/2021/5/e26387
doi:10.2196/26387
PMID:33970109

Letter to the Editor

And Justice for All? There Is More to the Interoperability of Contact Tracing Apps Than Legal Barriers. Comment on "COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control"

Rik Crutzen[1], PhD

Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Rik Crutzen, PhD
Department of Health Promotion
Care and Public Health Research Institute
Maastricht University
PO Box 616
Maastricht, 6200 MD
Netherlands
Phone: 31 433882828
Email: rik.crutzen@maastrichtuniversity.nl

Related Articles:

Comment on: https://mhealth.jmir.org/2020/11/e23194/

Comment in: https://mhealth.jmir.org/2021/5/e26630/

(JMIR Mhealth Uhealth 2021;9(5):e26218) doi:10.2196/26218

KEYWORDS

COVID-19; contact tracing; data protection; privacy; interoperability; global health; public health

Du and colleagues [1] have conducted a thorough analysis of legal barriers for realizing the interoperability of contact tracing apps and emphasize the need for developing coordinated solutions to promote safe international travel and global control of the COVID-19 pandemic. Possible options for destroying the technological Tower of Babel are proposed that, in my opinion, are not necessarily needed due to legal barriers and warrant a broader reflection.

First, Du et al [1] note that the region-based development of contact tracing apps, along with the data protection laws used in different countries and regions, has resulted in a disconnection between contact tracing apps. A plea is made for a broader consensus in the international community. Within the European Union (EU), interoperability guidelines for contact tracing apps were already adopted by consensus by the eHealth Network in May 2020 [2]. So far, 22 countries make use of a contact tracing app that is in principle interoperable through a federation gateway [3].

Du et al [1] are partly right in stating that this will only solve the problem within the EU, partly because the GDPR (General Data Protection Regulation) imposes strict limitations on countries outside the EU to process data transfers. In short, other countries need to meet the high standards set by the GDPR in terms of personal data protection (for *all* individuals within the EU regardless of citizenship, including refugees and tourists). In other words, the GDPR does not rule out data transfer *in principle*, but sets high standards [4]. This because the right to protection of personal data is part of the Charter of Fundamental Rights of the EU [5]. Article 8 of the Charter clearly indicates that "everyone has the right to protection of personal data concerning him or her."

Second, Du et al [1] offer a bold proposal: a common contact tracing app that is accepted by all countries and made mandatory for international travelers. Technically, the Google-Apple API (application programming interface) allows governments to work on developing their own contract tracing apps that are interoperable. Although this is not the same as the proposed app, it is a common interface that can be used across countries.

More importantly, even if all legal barriers are addressed when adopting such an interface on a global scale, making a contact tracing app mandatory is too bold a proposal. This is not (only) a legal concern, but (even more so) an ethical concern. Morley

and colleagues [6] have synthesized 16 questions concerning factors that should be satisfied in order for a contact tracing app to be ethical. An important factor dictates that downloading and installing such an app should be optional. People should also not be penalized for noncompliance. Morley et al [6] acknowledge that these questions are likely to generate disagreement in terms of satisfying and prioritizing factors.

So, I agree with Du et al [1] on the need for developing coordinated solutions, but this should not only be focused on addressing legal barriers. Instead, such solutions should make optimal use of readily available technology and take ethical concerns seriously. In a raging pandemic, it might be alluring to take an everything-but-the-kitchen-sink approach and focus solely on controlling COVID-19. However, especially in a crisis, this "is dangerous when it ignores the real costs, including serious and long-lasting harms to fundamental rights and freedoms" [6]. This is what makes them fundamental.

Conflicts of Interest

RC serves as chair of the Task Force Behavioral Sciences of the Dutch Ministry of Health, Welfare and Sport, which advises on the development, implementation, and evaluation of digital solutions that contribute to the control of COVID-19.

References

1. Du L, Raposo VL, Wang M. COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194. [doi: 10.2196/23194]
2. Interoperability guidelines for approved contact tracing mobile applications in the EU. eHealth Network. 2020. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/contacttracing_mobileapps_guidelines_en.pdf [accessed 2021-04-16]
3. Explanatory memorandum on temporary law notification application Covid-19. Dutch Ministry of Health, Welfare and Sport. 2020. URL: https://www.rijksoverheid.nl/documenten/publicaties/2020/11/20/nota-van-toelichting [accessed 2021-04-16]
4. Regulation (EU)/679 of the European Parliament and of the Council of 27 April on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. European Parliament and Council. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504 [accessed 2021-04-16]
5. Charter of Fundamental Rights of the European Union. European Union. 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT [accessed 2021-04-16]
6. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]

Abbreviations

API: application programming interface
EU: European Union
GDPR: General Data Protection Regulation

Edited by T Derrick, L Buis; submitted 02.12.20; peer-reviewed by S Du; accepted 12.04.21; published 26.05.21.

Please cite as:
Crutzen R
And Justice for All? There Is More to the Interoperability of Contact Tracing Apps Than Legal Barriers. Comment on "COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control"
JMIR Mhealth Uhealth 2021;9(5):e26218
URL: https://mhealth.jmir.org/2021/5/e26218
doi:10.2196/26218
PMID:33848974

Letter to the Editor

Authors' Reply to: And Justice for All? There Is More to the Interoperability of Contact Tracing Apps Than Legal Barriers. Comment on "COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control"

Li Du[1], MBBS, LLB, PhD; Vera Lúcia Raposo[1,2], BA, LLB, LLM, PhD; Meng Wang[1], LLB, LLM

[1]Faculty of Law, University of Macau, Macau, China

[2]Faculty of Law, University of Coimbra, Coimbra, Portugal

Corresponding Author:
Li Du, MBBS, LLB, PhD
Faculty of Law
University of Macau
Avenida da Universidade
Taipa
Macau, 999078
China
Phone: 853 88224733
Email: stephendu@um.edu.mo

Related Articles:

Comment on: https://mhealth.jmir.org/2021/5/e26218/

Comment on: https://mhealth.jmir.org/2020/11/e23194/

(JMIR Mhealth Uhealth 2021;9(5):e26630) doi:10.2196/26630

KEYWORDS

COVID-19; contact tracing; data protection; privacy; interoperability; global health; public health

We appreciate Crutzen's comments [1] on our analysis [2] of the technologic Tower of Babel, which analyzed the global functioning and interaction of COVID-19 contact tracing apps.

In his critique, Crutzen argued that it is not all about the law, but also about ethics. We agree with this proposition. Ethics and the principles of digital governance are essential to structure the digital world [3], especially during (and after) a pandemic [4].

We also agree with his observation that the eHealth Network has contributed to achieving the interoperability guidelines for contact tracing apps within the European Union (EU). However, the utility of this network is limited to Europe. Because member states share a basic framework regarding data protection (ie, the General Data Protection Regulation [GDPR]), unsolved legal conflicts between national laws on privacy are not likely to happen. Without such a common ground for data protection, the Google-Apple API (application programming interface) code can only help governments around the world to fast track the adoption of contact tracing app technology, but it may not solve the interoperability of contact tracing apps between countries (eg, between an EU member state and the United States). Moreover, European countries are more willing to work together in this regard and to adapt their national laws and technical practices, as it has already happened within several other domains: passports, border controls, currency, VAT (value-added tax) rates, and roaming. These are all areas related to the free circulation of people that EU member states have agreed to "harmonize" in order to facilitate free movement and a single European market [5]. Guidelines for interoperability between contact tracing mobile apps are just another step toward achieving that goal. On page 6, we underlined that "a broader consensus is required for the international community, since a common European approach will only solve the problem within the European Union."

In general, we are concerned with both the legal and ethical disconformity between different apps, while Crutzen focuses primarily on ethical concerns and the protection of human rights. In a sense, the two papers appear complementary, rather than to be in conflict with each other.

We take issue with one of Crutzen's statements, specifically the idea that making a contact tracing app mandatory for international travelers is unacceptable. In this case, we would need to abolish several other requisites imposed on international travelers, such as visa requirements for certain destinations or even the mere presentation of passports and other traveling documents. The most paradigmatic example is the proof of immunization required for some destinations, which, like the mandatory contact tracing app, is required for public health reasons and supported by institutions such as the World Health Organization [6]. Under Crutzen's reasoning, we would have to conclude that vaccination requirements are unethical and constitute a violation of travelers' rights.

Moreover, many countries are still requiring a 14-day compulsory quarantine for international travelers [7]. Using the contact tracing app for international travelers would actually make international travel easier, giving people more freedom for traveling during the pandemic. Even if the mandatory use of contact tracing apps can be considered burdensome, it would facilitate the prevention of more severe intrusions in travelers' personal freedom.

A contact tracing app that is mandatory for international travelers is a threat to fundamental rights. We recognize that. However, in light of the delicate balance between public health and individual rights and freedoms, and considering that from the various restrictions that can be imposed, we believe this measure is fully acceptable, especially based on an assessment of necessity, proportionality, and adequacy. It is certainly more acceptable than the ban on international traveling still in place in many parts of the world.

Conflicts of Interest

None declared.

References

1. Crutzen R. And Justice For All? There Is More To Interoperability of Contact Tracing Apps Than Legal Barriers. Comment on "COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control". JMIR Mhealth Uhealth 2021 May;9(5):e26218. [doi: 10.2196/26218] [Medline: 33848974]
2. Du L, Raposo VL, Wang M. Correction: COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194 [FREE Full text] [doi: 10.2196/26239] [Medline: 33275590]
3. Floridi L. Soft Ethics: Its Application to the General Data Protection Regulation and Its Dual Advantage. Philos. Technol 2018 Jun 3;31(2):163-167. [doi: 10.1007/s13347-018-0315-5]
4. Taddeo M. The Ethical Governance of the Digital During and After the COVID-19 Pandemic. Minds Mach (Dordr) 2020 Jun 12:1-6 [FREE Full text] [doi: 10.1007/s11023-020-09528-5] [Medline: 32836869]
5. Kaeding M. Towards an Effective European Single Market. Switzerland: Springer VS; 2013.
6. Yellow fever vaccination essential for Angola, WHO reminds travellers. World Health Organization. 2016 Apr 26. URL: https://www.who.int/news/item/26-04-2016-yellow-fever-vaccination-essential-for-angola-who-reminds-travellers [accessed 2020-12-18]
7. Schwartz K. I'm a U.S. Citizen. Where in the World Can I Go? New York Times. 2021 Apr 15. URL: https://www.nytimes.com/article/coronavirus-travel-restrictions.html [accessed 2021-04-16]

Abbreviations

API: application programming interface
EU: European Union
GDPR: General Data Protection Regulation
VAT: value-added tax

Edited by T Derrick, L Buis; submitted 19.12.20; this is a non–peer-reviewed article;accepted 12.04.21; published 26.05.21.

Please cite as:
Du L, Raposo VL, Wang M
Authors' Reply to: And Justice for All? There Is More to the Interoperability of Contact Tracing Apps Than Legal Barriers. Comment on "COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control"
JMIR Mhealth Uhealth 2021;9(5):e26630
URL: https://mhealth.jmir.org/2021/5/e26630
doi:10.2196/26630
PMID:33852409

Original Paper

A Mobile Phone–Based Intervention to Reduce Mental Health Problems in Health Care Workers During the COVID-19 Pandemic (PsyCovidApp): Randomized Controlled Trial

Maria Antònia Fiol-DeRoque[1,2*], PhD; Maria Jesús Serrano-Ripoll[1,2,3*], PhD; Rafael Jiménez[1,3], PhD; Rocío Zamanillo-Campos[1,2,4], PhD; Aina María Yáñez-Juan[1,5], PhD; Miquel Bennasar-Veny[1,6], PhD; Alfonso Leiva[1,2,7], PhD; Elena Gervilla[1,3], PhD; M Esther García-Buades[3], PhD; Mauro García-Toro[1], PhD, MD; Pablo Alonso-Coello[8,9], PhD, MD; Guadalupe Pastor-Moreno[8,10], PhD; Isabel Ruiz-Pérez[8,10], PhD, MD; Carolina Sitges[1,3,11], PhD; Javier García-Campayo[7,12], PhD, MD; Joan Llobera-Cánaves[1,2,7], PhD, MD; Ignacio Ricci-Cabello[1,2,8], PhD

[1]Health Research Institute of the Balearic Islands, Palma de Mallorca, Spain

[2]Primary Care Research Unit of Mallorca, Balearic Islands Health Services, Palma de Mallorca, Spain

[3]Department of Psychology, University of the Balearic Islands (UIB), Palma de Mallorca, Spain

[4]Department of Health, Valencian International University (VIU), Valencia, Spain

[5]Research Group on Global Health & Human Development, University of the Balearic Islands, Palma de Mallorca, Spain

[6]Department of Nursing and Physiotherapy, University of the Balearic Islands, Palma de Mallorca, Spain

[7]Primary Care Prevention and Health Promotion Research Network, Madrid, Spain

[8]CIBER de Epidemiología y Salud Pública, Madrid, Spain

[9]Department of Clinical Epidemiology and Public Health, Biomedical Research Institute Sant Pau, Iberoamerican Cochrane Centre, Barcelona, Spain

[10]Andalusian School of Public Health, Granada, Spain

[11]Department of Psychology, Research Institute of Health Sciences, Palma de Mallorca, Spain

[12]Aragon Institute for Health Research, Miguel Servet University Hospital, Zaragoza, Spain

[*]these authors contributed equally

Corresponding Author:
Maria Jesús Serrano-Ripoll, PhD
Health Research Institute of the Balearic Islands
Edificio S, Hospital Universitario Son Espases
Carretera de Valldemossa
Palma de Mallorca, 07120
Spain
Phone: 34 +34610753696
Email: mariajesus.serranoripoll@ssib.es

Abstract

Background: The global health emergency generated by the COVID-19 pandemic is posing an unprecedented challenge to health care workers, who are facing heavy workloads under psychologically difficult situations. Mental mobile Health (mHealth) interventions are now being widely deployed due to their attractive implementation features, despite the lack of evidence about their efficacy in this specific population and context.

Objective: The aim of this trial is to evaluate the effectiveness of a psychoeducational, mindfulness-based mHealth intervention to reduce mental health problems in health care workers during the COVID-19 pandemic.

Methods: We conducted a blinded, parallel-group, controlled trial in Spain. Health care workers providing face-to-face health care to patients with COVID-19 were randomly assigned (1:1) to receive the PsyCovidApp intervention (an app targeting emotional skills, healthy lifestyle behavior, burnout, and social support) or a control app (general recommendations about mental health care) for 2 weeks. The participants were blinded to their group allocation. Data were collected telephonically at baseline and after 2 weeks by trained health psychologists. The primary outcome was a composite of depression, anxiety, and stress (overall score on the Depression Anxiety Stress Scale-21 [DASS-21]). Secondary outcomes were insomnia (Insomnia Severity Index), burnout (Maslach Burnout Inventory Human Services Survey), posttraumatic stress (Davidson Trauma Scale), self-efficacy (General

Self-Efficacy Scale), and DASS-21 individual scale scores. Differences between groups were analyzed using general linear modeling according to an intention-to-treat protocol. Additionally, we measured the usability of the PsyCovidApp (System Usability Scale). The outcome data collectors and trial statisticians were unaware of the treatment allocation.

Results: Between May 14 and July 25, 2020, 482 health care workers were recruited and randomly assigned to PsyCovidApp (n=248) or the control app (n=234). At 2 weeks, complete outcome data were available for 436/482 participants (90.5%). No significant differences were observed between the groups at 2 weeks in the primary outcome (standardized mean difference –0.04; 95% CI –0.11 to 0.04; *P*=.15) or in the other outcomes. In our prespecified subgroup analyses, we observed significant improvements among health care workers consuming psychotropic medications (n=79) in the primary outcome (–0.29; 95% CI –0.48 to –0.09; *P*=.004), and in posttraumatic stress, insomnia, anxiety, and stress. Similarly, among health care workers receiving psychotherapy (n=43), we observed improvements in the primary outcome (–0.25; 95% CI –0.49 to –0.02; *P*=.02), and in insomnia, anxiety, and stress. The mean usability score of PsyCovidApp was high (87.21/100, SD 12.65). After the trial, 208/221 participants in the intervention group (94.1%) asked to regain access to PsyCovidApp, indicating high acceptability.

Conclusions: In health care workers assisting patients with COVID-19 in Spain, PsyCovidApp, compared with a control app, reduced mental health problems at 2 weeks only among health care workers receiving psychotherapy or psychotropic medications.

Trial Registration: ClinicalTrials.gov NCT04393818; https://clinicaltrials.gov/ct2/show/NCT04393818.

(JMIR Mhealth Uhealth 2021;9(5):e27039) doi:10.2196/27039

KEYWORDS

COVID-19; randomized controlled trial; mental health; health care workers; mHealth; app

Introduction

Since the declaration of the COVID-19 pandemic, the disease has spread globally, with almost 55 million known cases and a death toll of over 1.3 million people [1]. Early anecdotal evidence from Wuhan showed how this unprecedented situation impacted the mental health of frontline health care workers [2]. This was confirmed by subsequent systematic reviews [3,4], in which a remarkably high prevalence of acute stress (40%), anxiety (30%), burnout (28%), depression (24%), and posttraumatic stress disorder (13%) was observed among frontline health care workers. Working in an environment with a high risk of infection, job-related stress, heavy workloads, and lack of protective equipment were found to significantly exacerbate these psychological problems [3]. On May 8, 2020, Spain reported the highest cumulative number of COVID-19 infections among health care workers worldwide (30,663 infections, accounting for 20% of all infections in health care workers worldwide) [5]. Not surprisingly, the psychological consequences disproportionally affected the mental health of Spanish health care workers, with approximately 57% of them presenting symptoms of posttraumatic stress disorder, 59% presenting symptoms of anxiety disorder, 46% presenting symptoms of depressive disorder, and 41% feeling emotionally drained [6].

Health services worldwide are being urged to implement strategies to mitigate the severe psychological consequences experienced by health care workers. Among the different types of strategies considered, mobile health (mHealth) interventions are receiving special attention [7] not only because of their attractive implementation features but also because they can be delivered in the absence of face-to-face interactions, reducing the risk of infection with SARS-CoV-2. Further, they can address non–treatment-seeking behavior (a common issue among health care workers [8]), as they provide the opportunity to engage individuals in need of treatment in a timely and anonymous fashion. This growing interest in mHealth interventions is supported by the positive results of acceptance rates [9] and sustainability [10] observed in other contexts and populations. Recent trials have examined the efficacy of mHealth interventions addressing mental health problems, including depression [11,12], suicide [13], schizophrenia [14], substance use disorders [15], and psychosis [16], among others [17]. Recent systematic reviews investigating the efficacy of smartphone apps for mental health show that these apps can produce significant reductions in anxiety [18] and depression [19]. However, the effectiveness of mHealth interventions in health care workers in the COVID-19 pandemic context is largely unknown, as observed by recent reviews that highlighted the lack of evaluations of client-relevant outcomes [20] and the lack of both quantitative and qualitative evidence to inform the selection of interventions that are beneficial to the mental health of frontline health care workers [21]. Hence, robust, large-scale trials are urgently needed to determine the extent to which mHealth interventions can improve the mental health of frontline health care workers.

This blinded, individually randomized, parallel-group, controlled trial aimed to evaluate the effectiveness of PsyCovidApp (a self-managed and self-guided psychoeducational mobile-based intervention with no therapist support) to reduce symptoms of depression, anxiety, stress and other mental health problems in health care workers during the COVID-19 pandemic in Spain.

Methods

Design and Setting

We conducted a blinded, individually randomized, parallel-group, controlled trial in Spain. Because the ultimate goal of the study was to inform decisions about rolling up the intervention to make it available to all health care workers in Spain, we used a pragmatic approach. Pragmatic trials are ideally suited to inform choices between treatments because as opposed to exploratory trials (which typically examine treatment benefits under ideal conditions using carefully defined subjects),

pragmatic trials enable measurement of the effectiveness of interventions under real conditions in a sample of participants to whom the treatment would be applied [22]. Participants were individually randomized with an allocation ratio of 1:1 to receive either the PsyCovidApp intervention or the control app over 2 weeks. Ethical approval was obtained by the Research Ethics Committee of the Balearic Islands (IB 4216/20 PI). The study protocol and statistical analysis plan have been published previously (ClinicalTrials.gov NCT04393818) [23].

Participants

The target population was male and female health care workers aged >18 years who had provided health care to patients with COVID-19 during the viral outbreak in Spain. We included health care workers from any medical specialty (pneumology, internal medicine, emergency, primary care, etc) and role (physicians, nurses, nurse assistants, etc) with access to a smartphone. We included health care workers who had provided direct, face-to-face health care to patients with a diagnosis of infection with COVID-19. We excluded health care workers who were not able to download and activate the app used to deliver the intervention during the next 10 days following the baseline assessment.

Following a purposive sampling method, we sent invitations to health care workers to participate in the trial through direct contact via email and telephone to key stakeholders (115 hospital and home care centers, 138 professional associations, and 8 scientific societies and trade unions) and by social media. Health care workers who were willing to participate registered their interest by completing a web-based questionnaire, consenting to be contacted telephonically. A team of 23 health psychologists who had previously received a 2-hour training session (to ensure homogeneity in recruitment, questionnaire administration, and data entry methods) contacted the registered health care workers by telephone to confirm the eligibility criteria and obtain informed consent (audio-recorded). The recruitment period spanned 10 weeks, from May 14 to July 25, 2020. Participants were enrolled on a rolling “first-come-first-served” basis until target sample sizes were met. To incentivize trial participation, we offered participation certificates to all health care workers completing the postintervention assessment.

Randomization and Masking

Participants were randomly assigned (1:1) to receive the PsyCovidApp intervention or the control app over 2 weeks by a designated researcher (MAF, who was not involved in data collection or analysis) using a computer-generated sequence of random numbers created by internet relay chat (IRC). Randomization was not stratified. Health care workers were blinded to group allocation (as both groups received an app). The outcome data collectors and trial statisticians were unaware of the treatment allocation.

Procedures

Immediately after obtaining informed consent, a team of psychologists conducted a psychological (preintervention) evaluation via telephone interview and instructed participants on how to download the Clinicovery App (Apploading, Inc). Clinicovery is the app that was used to deliver either the contents of the PsyCovidApp intervention or the control contents. Within 48 hours after participants successfully downloaded and activated the app (user activation was used as a checkpoint to ensure the participants could successfully access the app), a member of our research team loaded the contents to the app according to the group the participants had been allocated to. During the following 14 days, all health care workers had access to the content of their assigned group (PsyCovidApp intervention or control). The PsyCovidApp intervention was developed by a group of psychologists (MJSR, EG, CS, RJ, MEGB), psychiatrists (JGC, MGT), and experts in healthy lifestyle promotion (AMYJ, MBV), informed by findings from an exploratory qualitative study involving in-depth interviews with 9 health care workers seeking psychological support as a result of their professional activity during the COVID-19 pandemic (unpublished results). PsyCovidApp was specifically designed to prevent and mitigate the most frequent mental problems suffered by health care workers who are dealing with the COVID-19 emergency (depression, anxiety, posttraumatic stress, and burnout). A detailed description of the intervention is available elsewhere [23]. In short, the self-managed psychoeducational intervention, based on cognitive-behavioral therapy and mindfulness approaches, included written and audiovisual content targeting four areas: emotional skills, healthy lifestyle behavior, work stress and burnout, and social support. Additionally, the intervention included daily prompts (notifications) that included brief questionnaires to monitor mental health status, followed by short messages offering tailored information and resources based on the participants´ responses. The full content of the intervention is available in Multimedia Appendix 1.

Participants in the Control App group had access through the Clinicovery app to brief written information about the mental health care of health care workers during the COVID-19 pandemic (adapted from a set of materials developed by the Spanish Society of Psychiatry; the contents are available in Multimedia Appendix 2).

After 2 weeks, the apps in both groups were disabled, and a postintervention psychological assessment was undertaken. The follow-up was undertaken via telephone between 24 hours and 10 days after the intervention concluded, and it included the same questionnaires used in the first evaluation and the System Usability Scale (SUS) [24]. Once the postintervention assessment had finished, all participants were offered free, unrestricted access to PsyCovidApp.

Outcome Measures

The primary outcome was an overall index of depression, anxiety, and stress (overall score of the Spanish version of the Depression, Anxiety, and Stress Scale [DASS-21] instrument [25]) assessed at 2 weeks. The score ranges from 0 (best outcome) to 21 (worst outcome). The instrument contains three 7-item subscales assessing the presence and intensity of depression, anxiety, and stress. The items are based on a Likert scale ranging from 0 to 3 points. The instrument shows adequate internal consistency (Cronbach α=.91) and construct validity (3-factor structure identified after exploratory factor analysis, explaining 50% of the total variance) [26].

Secondary outcome measures were the difference between the intervention and control groups in the mean scores of the following instruments:

- Davidson Trauma Scale (DTS) [27]. The DTS is a 17-item self-report Likert scale instrument that assesses the 17 symptoms of posttraumatic stress disorder in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Both a frequency and a severity score can be determined. The DTS yields a frequency score (ranging from 0 to 68), severity score (ranging from 0 to 68), and total score (ranging from 0 to 136). Higher scores are indicative of a worse outcome. The Spanish version of the instrument shows adequate reliability (test-retest intraclass correlation=0.87; α=.90) as well as high discriminant validity [28]. For this study, the instrument was adapted to measure posttraumatic stress disorder since the onset of the COVID-19 pandemic. This adaptation consisted of reformulating the stem of all the items, replacing "during the last week" with "since the onset of the current COVID-19 pandemic".
- Maslach Burnout Inventory - Human Services Survey (MBI-HSS) [29], which yields specific scores for each of its three subscales (emotional exhaustion, α=.85; depersonalization, α=.58; and professional accomplishment, α=.71). The Spanish version of the instrument shows adequate internal consistency (except for the depersonalization subscale) and adequate factorial validity [30].
- Insomnia Severity Index (ISI) [31]. The ISI is a 7-item, self-reported Likert scale instrument assessing the severity of both nighttime and daytime components of insomnia. Scores range from 0 (best outcome) to 28 (worst outcome). In the Spanish version, principal component analysis showed only one factor explaining 69% of the total variance, with an internal consistency reliability of 0.91. Regarding its validity, the ISI shows statistically significant positive correlations with the Athens Insomnia Scale-5 (r=0.93) and negative correlations with the Mini Mental State Examination (r=–0.15) [32].
- General Self-Efficacy Scale (GSE) [33]. The GSE is a 10-item, Likert-scale, self-reported instrument that assesses optimistic self-beliefs to cope with a variety of difficult demands in life. Scores range from 10 (worst outcome) to 40 (best outcome). The Spanish version shows adequate internal consistency (α=.85) and construct validity [34].
- Individual subscales of the DASS-21 instrument: depression (α=.90), anxiety (α=.88), and stress (α=.88) [35,36].
- Usability of PsyCovidApp at postintervention: SUS [24]. Higher scores are indicative of higher usability, and scores above the 68-point threshold can be considered to indicate high usability. The Spanish version shows adequate internal consistency (α=.80) and concurrent validity (significant correlation with the Adjective Rating Scale; r=0.56) [37].

Statistical Analysis

The sample size and power calculations have been described previously [23]. We estimated that 440 participants (220 per group, allowing for 10% attrition) would be required to detect at least an effect size of 0.25 (standardized between-group mean difference) on DASS-21 with 80% power and 5% α (one-sided).

The analyses followed the agreed statistical analysis plan, published before database lock [23]. Descriptive statistics summarizing prerandomization variables, and outcome measures at 2 weeks were reported by treatment group and overall. Differences between groups of primary and secondary outcomes were analyzed using general linear modelling (analysis of covariance) for continuous variables, adjusted by baseline score. We report standardized between-group differences in primary and secondary outcome measures at 2 weeks. In those outcomes that were interpreted as being in favor of PsyCovidApp if the standardized group difference was more than 0, the estimated effect was reflected (ie, multiplied by –1) so that outcomes for which the standardized group difference was lower than 0 could be homogenously interpreted as being in favor of the intervention. In the primary statistical analysis, all health care workers who agreed to participate were included in the analysis according to the group to which they were assigned. We used multiple imputation by chained equations to fill in missing values (50 imputation sets) [38]. In Multimedia Appendix 3, we report unstandardized between-group differences. In the analysis, *P* values and CIs were not corrected for multiple secondary outcome comparisons.

We conducted three prespecified subgroup analyses to examine the impact of the PsyCovidApp intervention on the primary and secondary outcomes based on the following baseline characteristics: use of psychotropic medications (yes vs no), use of psychotherapy (yes vs no), and symptomatology of depression, anxiety, and stress (yes vs no, based on baseline DASS-21 median overall score). We conducted statistical tests for interaction (including an interaction term in the models) to determine whether chance was an unlikely explanation for the apparent subgroup effects identified. We used the Instrument to Assess the Credibility of Effect Modification Analyses (ICEMAN) [39] to assess the credibility of our subgroup analyses. As a sensitivity analysis, we reanalyzed all outcomes on a complete case basis (ie, without imputation or adjustment for baseline predictors of missingness). We used SPSS, version 25 (IBM Corporation) and Stata, version 13 (StataCorp LLC) to conduct the statistical analyses.

Role of the Funding Source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Between May 14 and July 25, 2020, 684 health care workers submitted an expression of interest in enrolling in the PsyCovidApp trial. 482 eligible participants provided informed consent and were randomly assigned to the PsyCovidApp intervention group (n=248) or the Control App group (n=234; Figure 1). Recruitment by region is shown in Multimedia

Appendix 4. Participants included health care workers from all regions in Spain except the Canary Islands and Cantabria.

The baseline characteristics were balanced between the groups (Table 1). Most participants were women (401/482; 83.2%), and the median age was 42 years (IQR 33-49). Approximately one-third (161, 33.4%) of the 482 participants were nurses, 153 (31.7%) were physicians, and 147 (30.5%) were nurse assistants. Most worked in the hospital setting: 98/482 (20.3%) in internal medicine, 81/482 (16.8%) in intensive care units, 79/482 (16.4%) in hospital emergency units, 31/482 (6.4%) in infection units, and 103/482 (21.4%) in other hospital units. Of the 482 participants, 61 (12.7%) worked in primary care and 29 (6%) in home-care settings.

Figure 1. Trial profile. Multiple imputation was used to facilitate the overall sample analysis; all randomized participants contributed to the statistical analysis.

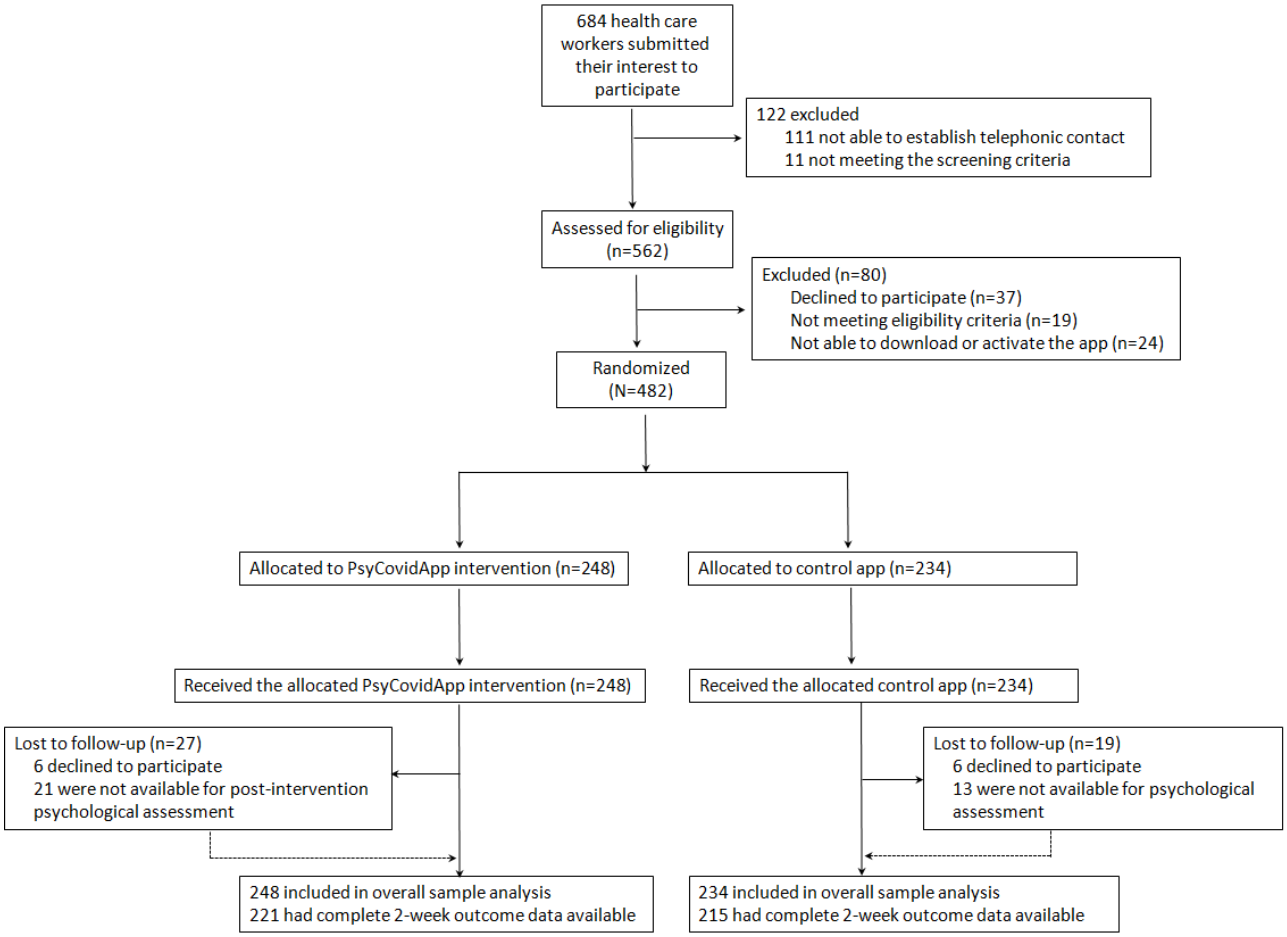


Table 1. Baseline demographic characteristics of the intention-to-treat population (N=482).

Characteristic	PsyCovidApp group (n=248)	Control App group (n=234)	Total (N=482)
Age, years			
Mean (SD)	42.07 (11.0)	40.62 (9.6)	41.37 (10.4)
Median (IQR; range)	42 (34-51; 23-63)	41 (32-47; 23-61)	41.5 (33-49; 23-63)
<36, n (%)	75 (30.2)	82 (35)	157 (32.6)
36-45, n (%)	79 (31.9)	81 (34.6)	160 (33.2)
46-55, n (%)	60 (24.2)	52 (22.2)	112 (23.2)
>55, n (%)	34 (13.7)	19 (8.1)	53 (11)
Gender, n (%)			
Male	38 (15.3)	43 (18.4)	81 (16.8)
Female	210 (84.7)	191 (81.6)	401 (83.2)
Occupational role, n (%)			
Physician	76 (30.6)	77 (32.9)	153 (31.7)
Nurse	87 (35.1)	74 (31.6)	161 (33.4)
Nurse assistant	77 (31)	70 (29.9)	147 (30.5)
Other	8 (3.2)	13 (5.6)	22 (4.6)
Setting, n (%)			
Primary care	35 (14.1)	26 (11.1)	61 (12.7)
Internal medicine	48 (19.4)	50 (21.4)	98 (20.3)
Intensive care unit	40 (16.1)	41 (17.5)	81 (16.8)
Hospital emergencies unit	31 (12.5)	48 (20.5)	79 (16.4)
Home care	19 (7.7)	10 (4.3)	29 (6)
Infection unit	16 (6.5)	15 (6.4)	31 (6.4)
Other hospital unit	59 (23.8)	44 (18.8)	103 (21.4)
Time working with patients with COVID-19 (weeks), n (%)			
<2	9 (3.6)	4 (1.7)	13 (2.7)
2-4	20 (8.1)	14 (6)	34 (7.1)
>4	219 (88.3)	216 (92.3)	435 (90.2)
Infected with COVID-19, n (%)			
Yes	31 (12.5)	34 (14.5)	65 (13.5)
No	214 (86.3)	195 (83.3)	409 (84.9)
Unknown	3 (1.2)	5 (2.1)	8 (1.7)
Perception about the adequacy of available measures to protect health care workers from COVID-19, n (%)			
Inadequate measures	84 (33.9)	77 (32.9)	161 (33.4)
Adequate measures	163 (65.7)	157 (67.1)	320 (66.4)
Perception about the information about the procedures to provide health care to patients with COVID-19, n (%)			
Inadequate information	101 (40.7)	116 (49.6)	217 (45)
Adequate information	147 (59.3)	117 (50)	264 (54.8)
Currently consuming psychotropic medication, n (%)			
No	207 (83.5)	196 (83.8)	403 (83.6)
Yes	41 (16.5)	38 (16.2)	79 (16.4)
Currently receiving psychotherapy, n (%)			
No	227 (91.5)	212 (90.6)	439 (91.1)

Characteristic	PsyCovidApp group (n=248)	Control App group (n=234)	Total (N=482)
Yes	21 (8.5)	22 (9.4)	43 (8.9)

By the time of recruitment, most participants (435/482, 90.2%) had been providing health care to patients with COVID-19 for more than 4 weeks, and 65/482 (13.5%) had received a diagnosis of COVID-19 infection. Approximately one-third (161/482, 33.4%) perceived that the measures offered to protect them from COVID-19 had been inadequate, and 217/482 (45%) perceived that they had received inadequate information about the procedures to provide health care to patients with COVID-19. Of the 482 participants, 79 (16.4%) were using psychotropic medications, and 43 (8.9%) were receiving psychotherapy.

In relation to their mental health, 206 of the 482 participants (42.7%) presented symptoms of depression, 250 (51.9%) had symptoms of anxiety, 292 (60.6%) had symptoms of stress, 194 (40.2%) had symptoms of posttraumatic stress, and 128 (26.6%) had symptoms of insomnia (Table 2). Concerning burnout, 282/482 participants (58.5%) presented emotional exhaustion, 165/482 (34.2%) presented emotional depersonalization, and 203/482 (42.1%) presented moderate or low professional accomplishment. The mean self-efficacy score was 32.2 out of 40 (SD 4.7), indicating a high level of self-efficacy.

Table 2. Baseline clinical characteristics of the intention-to-treat population (N=482).

Characteristic	PsyCovidApp group (n=248)	Control App group (n=234)	Total (N=482)
Depression, anxiety, and stress			
DASS-21[a] overall score, mean (SD)	5.8 (3.9)	6.1 (3.8)	6.0 (3.8)
Depression (DASS-21 subscale), n (%)			
No symptoms (<5 points)	143 (57.7)	133 (56.8)	276 (57.3)
Mild (5-6 points)	26 (10.5)	32 (13.7)	58 (12)
Moderate (7-10 points)	55 (22.2)	48 (20.5)	103 (21.4)
Severe (11-13 points)	15 (6)	11 (4.7)	26 (5.4)
Extremely severe (>13 points)	9 (3.6)	10 (4.3)	19 (3.9)
Anxiety (DASS-21 subscale), n (%)			
No symptoms (<4 points)	121 (48.8)	111 (47.4)	232 (48.1)
Mild (4 points)	32 (12.9)	25 (10.7)	57 (11.8)
Moderate (5-7 points)	40 (16.1)	42 (17.9)	82 (17)
Severe (8-9 points)	24 (9.7)	20 (8.5)	44 (9.1)
Extremely severe (>9 points)	31 (12.5)	36 (15.4)	67 (13.9)
Stress (DASS-21 subscale), n (%)			
No symptoms (<8 points)	104 (41.9)	86 (36.8)	190 (39.4)
Mild (8-9 points)	28 (11.3)	32 (13.7)	60 (12.4)
Moderate (10-12 points)	56 (22.6)	58 (24.8)	114 (23.7)
Severe (13-16 points)	43 (17.3)	45 (19.2)	88 (18.3)
Extremely severe (>16 points)	17 (6.9)	13 (5.6)	30 (6.2)
Posttraumatic stress (DTS[b]), n (%)			
No (<40 points)	150 (60.5)	138 (59)	288 (59.8)
Yes (≥40 points)	98 (39.5)	96 (41)	194 (40.2)
Burnout (MBI-HSS[c]), n (%)			
Emotional exhaustion			
Low (0-16 points)	95 (38.3)	89 (38)	184 (38.2)
Moderate (17-26 points)	61 (24.6)	50 (21.4)	111 (23)
High (>27 points)	92 (37.1)	95 (40.6)	187 (38.8)
Professional accomplishment subscale			
High (>39 points)	144 (58.1)	135 (57.8)	279 (57.9)
Moderate (32-38 points)	65 (26.2)	54 (23.1)	119 (24.7)
Low (0-31 points)	39 (15.7)	45 (19.2)	84 (17.4)
Depersonalization subscale			
Low (0-6 points)	163 (65.7)	154 (65.8)	317 (65.8)
Moderate (17-12 points)	50 (20.2)	31 (13.2)	81 (16.8)
High (>13 points)	35 (14.1)	49 (20.9)	84 (17.4)
Insomnia (ISI[d]), n (%)			
Not clinically significant (0-7 points)	102 (41.1)	87 (37.2)	189 (39.2)
Subthreshold insomnia (8-14 points)	89 (35.9)	76 (32.5)	165 (34.2)
Clinical insomnia (moderate severity) (15-21 points)	49 (19.8)	61 (26.1)	110 (22.8)

Characteristic	PsyCovidApp group (n=248)	Control App group (n=234)	Total (N=482)
Clinical insomnia (severe) (22–28)	8 (3.2)	10 (4.3)	18 (3.7)
Self-efficacy (GSE[e])			
Mean score out of 40 points (SD)	32.4 (4.7)	32 (4.7)	32 (4.7)

[a]DASS-21: Depression, Anxiety, and Stress Scale-21.

[b]DTS, Davidson Trauma Scale.

[c]MBI-HSS: Maslach Burnout Inventory - Human Services Survey.

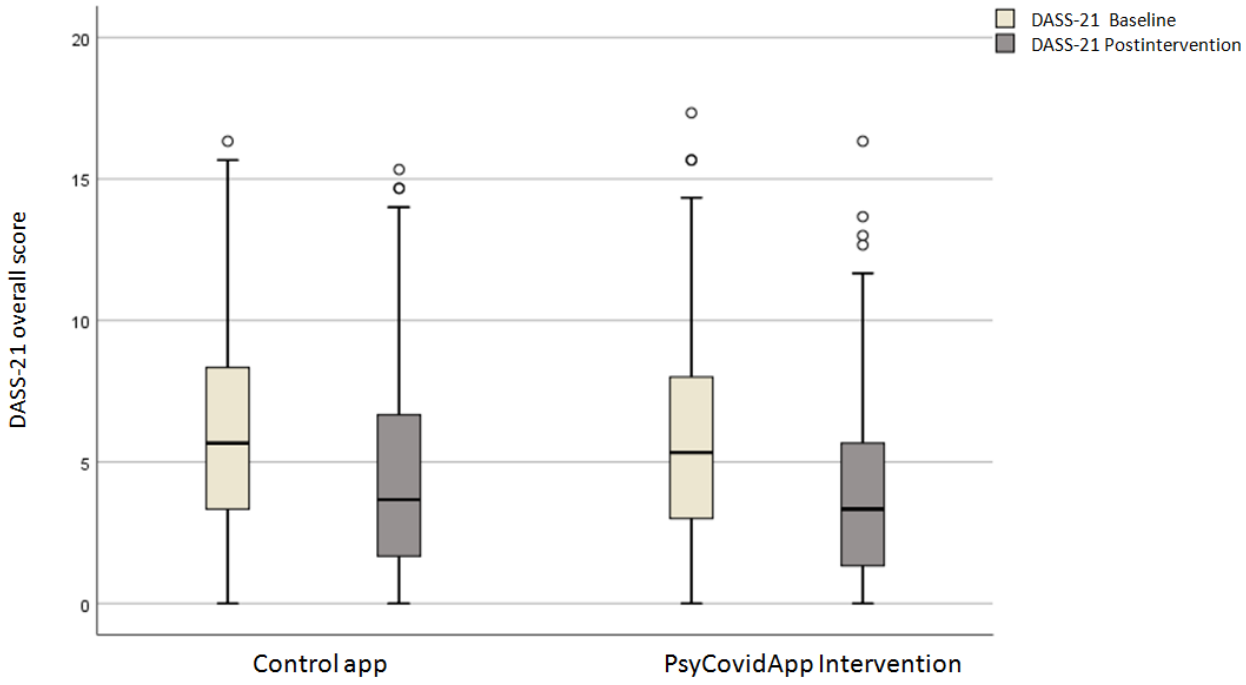
[d]ISI: Insomnia Severity Index.

[e]GSE: General Self-Efficacy Scale.

At 2 weeks, 27 of the 248 participants (10.9%) in the PsyCovidApp group and 19 of the 234 participants (8.1%) in the Control App group were lost to follow-up because they decided to withdraw from the study at the time of the postintervention psychological assessment (6 in the intervention group and 6 in the control group) or because we were unable to reach them for the telephonic postintervention psychological assessment (21 in the intervention group and 13 in the control group). None of the participants were deemed to be associated with reported adverse events or death.

Primary and secondary outcome data were available for 436 of the 482 participants (90.5%): 221 of 248 (89.1%) health care workers in the PsyCovidApp intervention group versus 215 of 234 (91.9%) in the Control App group. For the primary outcome, scale scores were lower at 2 weeks than at baseline in the PsyCovidApp and Control App groups (Figure 2). Similar reductions were observed at 2 weeks in both groups for all the secondary outcomes except for self-efficacy and depersonalization.

Figure 2. Changes in median DASS-21 scores over time, with the raw data plot of the median DASS-21 scores. Baseline scores were recorded before randomization.



The effect sizes for all outcomes are shown in Table 3, Table 4, and Multimedia Appendix 5. For the primary outcome, no significant differences in the DASS-21 overall score were identified between the groups at 2 weeks (standardized mean difference –0.04; 95% CI –0.11 to 0.04; *P*=.15). Similarly, none of the secondary outcomes significantly differed between groups at 2 weeks (all *P*>.05).

Table 3. Descriptive summaries of the primary and secondary outcome measures at baseline and 2 weeks for the PsyCovidApp and Control App groups.

Measure	At baseline						At 2 weeks					
	PsyCovidApp group (n=248)		Control App group (n=234)		Overall (N=482)		PsyCovidApp group (n=221)		Control App group (n=215)		Completers at follow-up (n=436)	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Primary outcome												
DASS-21[a] overall score	5.84 (3.85)	0-17.33	6.14 (3.77)	0-16.33	5.99 (3.81)	0-17.33	3.83 (3.21)	0-16.33	4.27 (3.47)	0-15.33	4.05 (3.35)	0-16.33
Secondary outcomes												
DASS-21 depression subscale	4.46 (4.13)	0-18	4.58 (4.02)	0-15	4.51 (4.07)	0-18	2.97 (3.49)	0-17	3.05 (3.65)	0-18	3.01 (3.56)	0-18
DASS-21 anxiety subscale	4.35 (3.86)	0-16	4.70 (4.25)	0-18	4.52 (4.06)	0-18	2.21 (2.43)	0-15	2.84 (3.36)	0-17	2.64 (3.13)	0-17
DASS-21 stress subscale	8.75 (5.07)	0-21	9.15 (4.63)	0-19	8.94 (4.86)	0-21	6.11 (4.50)	0-21	6.94 (4.68)	0-21	6.51 (4.60)	0-21
Posttraumatic stress (DTS[b])	34.57 (23.47)	0-117	36.91 (23.18)	0-100	35.71 (23.33)	0-117	24.91 (20.41)	0-96	26.36 (21.02)	0-91	25.62 (20.70)	0-96
Burnout (MBI-HSS[c]) emotional exhaustion subscale	23.27 (12.20)	0-54	23.57 (12.34)	0-54	23.41 (12.26)	0-54	19.43 (12.25)	0-51	19.67 (12.91)	0-54	19.54 (12.57)	0-54
Burnout (MBI-HSS) professional accomplishment subscale[d]	39.69 (6.43)	10-48	39.59 (6.62)	15-48	39.64 (6.52)	10-48	40.33 (6.31)	13-48	39.54 (6.93)	15-48	39.94 (6.63)	13-48
Burnout (MBI-HSS) depersonalization subscale	4.69 (5.08)	0-29	5.24 (5.41)	0-23	4.95 (5.25)	0-29	4.51 (4.96)	0-29	4.78 (5.25)	0-23	4.64 (5.10)	0-29
Insomnia (ISI[e])	9.80 (6.19)	0-26	10.16 (6.53)	0-27	9.98 (6.36)	0-27	8.07 (6.18)	0-28	8.44 (6.68)	0-23	8.25 (6.43)	0-28
Self-efficacy (GSE[f])[d]	32.42 (4.71)	19-40	32.00 (4.73)	18-40	32.21 (4.72)	18-40	33.22 (4.65)	18-40	32.54 (4.88)	17-40	32.88 (4.77)	17-40

[a]DASS-21: Depression, Anxiety, and Stress Scale-21.

[b]DTS: Davidson Trauma Scale.

[c]MBI-HSS: Maslach Burnout Inventory - Human Services Survey.

[d]Scale scores are reversed to homogeneously convey a similar treatment effect.

[e]ISI: Insomnia Severity Index.

[f]GSE: General Self-Efficacy Scale.

Table 4. Comparison of outcome measures between the PsyCovidApp and the Control App groups at 2 weeks. Data are adjusted standardized between-group mean differences with 95% CIs in parentheses. *P* values are not adjusted for multiple testing.

Measure	Sample of completers at follow-up (n=436)		Overall sample[a] (N=482)	
	Adjusted standardized between-group mean differences (95% CI)	*P* value	Adjusted standardized between-group mean differences (95% CI)	*P* value
DASS-21[b] overall score	–0.04 (-0.11 to 0.04)	.16	–0.04 (–0.11 to 0.04)	.15
DASS-21 depression subscale	0.00 (–0.07 to 0.08)	.47	0.00 (–0.07 to 0.08)	.47
DASS-21 anxiety subscale	–0.04 (–0.12 to 0.04)	.15	–0.04 (–0.12 to 0.04)	.17
DASS-21 stress subscale	–0.06 (–0.14 to 0.01)	.06	–0.06 (–0.14 to 0.01)	.05
Posttraumatic stress (DTS[c])	0.00 (–0.06 to 0.06)	.47	0.00 (–0.06 to 0.07)	.47
Burnout (MBI-HSS[d]) emotional exhaustion subscale	0.01 (–0.06 to 0.08)	.38	0.01 (–0.06 to 0.08)	.39
Burnout (MBI-HSS) professional accomplishment subscale[e]	–0.04 (–0.12 to 0.03)	.13	–0.05 (–0.12 to 0.03)	.12
Burnout (MBI-HSS) depersonalization subscale	0.01 (–0.06 to 0.09)	.36	0.01 (–0.06 to 0.09)	.36
Insomnia (ISI[f])	0.01 (–0.05 to 0.07)	.38	0.01 (–0.05 to 0.07)	.38
Self-efficacy (GSE[g])[e]	–0.02 (–0.10 to 0.05)	.26	–0.02 (–0.01 to 0.05)	.27

[a]Overall sample, derived by multiple imputation (50 imputations).

[b]DASS-21: Depression, Anxiety, and Stress Scale-21.

[c]DTS: Davidson Trauma Scale.

[d]MBI-HSS, Maslach Burnout Inventory - Human Services Survey.

[e]Scale scores reversed to homogeneously convey a similar treatment effect.

[f]ISI: Insomnia Severity Index.

[g]GSE: General Self-Efficacy Scale.

The impact of the intervention on prespecified subgroups of health care workers is presented in Figures 3 and 4 and Multimedia Appendix 3. In the subgroup of health care workers consuming psychotropic medications (n=79) (Figure 3), the PsyCovidApp group presented significantly lower DASS-21 overall scores (suggesting improved mental health) at 2 weeks than the Control App group (adjusted standardized mean difference –0.29; 95% CI –0.48 to –0.09; *P*=.004). Compared to the control app, the PsyCovidApp intervention significantly improved symptoms of anxiety (–0.26; 95% CI –0.45 to –0.08; *P*=.004), stress (–0.30; 95% CI –0.50 to –0.09; *P*=.003), posttraumatic stress (–0.20; 95% CI –0.37 to –0.03; *P*=.01), and insomnia (–0.16; 95% CI –0.30 to –0.02; *P*=.01); meanwhile, no differences were observed for symptoms of depression, emotional exhaustion, professional accomplishment, depersonalization, or self-efficacy (all *P*>.05). No significant differences were observed in any of the outcomes in the group of health care workers not consuming psychotropic medications (n=403). The interaction *P* values for anxiety, stress, posttraumatic stress, and insomnia were all <.05, suggesting that the apparent interaction was not a chance finding.

In the subgroup of participants receiving psychotherapy (n=43) (Figure 4), the PsyCovidApp group presented significantly lower DASS-21 overall scores (suggesting improved mental health) at 2 weeks than the Control App group; adjusted standardized mean difference –0.25; 95% CI –0.49 to –0.02; *P*=.02). Compared to the control app, the PsyCovidApp intervention significantly improved symptoms of anxiety (–0.24; 95% CI –0.48 to 0.00; *P*=.02), stress (–0.27; 95% CI –0.55 to 0.001; *P*=.02), and insomnia (–0.20; 95% CI –0.42 to 0.02; *P*=.03); meanwhile. no statistically significant differences were observed for symptoms of depression, posttraumatic stress disorder, emotional exhaustion, professional accomplishment, depersonalization, or self-efficacy (*P*>.05). No statistically significant differences were observed in any of the outcomes in the group of health care workers not receiving psychotherapy (n=439). The interaction *P* values for anxiety, stress, and insomnia were <.05.

No statistically significant differences (*P*>.05) were observed in the primary outcome or in any of the secondary outcomes examined in the subgroups of health care workers with higher and lower baseline DASS-21 scores (based on baseline DASS-21 median overall score).

Figure 3. Standardized mean differences for primary and secondary outcomes in healthcare workers reporting the use of psychotropic medications at baseline. Forest plot of standardized group differences between PsyCovidApp and Control App groups for all outcomes, whereby an effect lower than 0 favored the PsyCovidApp group. Error bars show 95% Confidence Intervals (CIs). DASS-21, Depression, Anxiety, and Stress Scale. *P<.05.

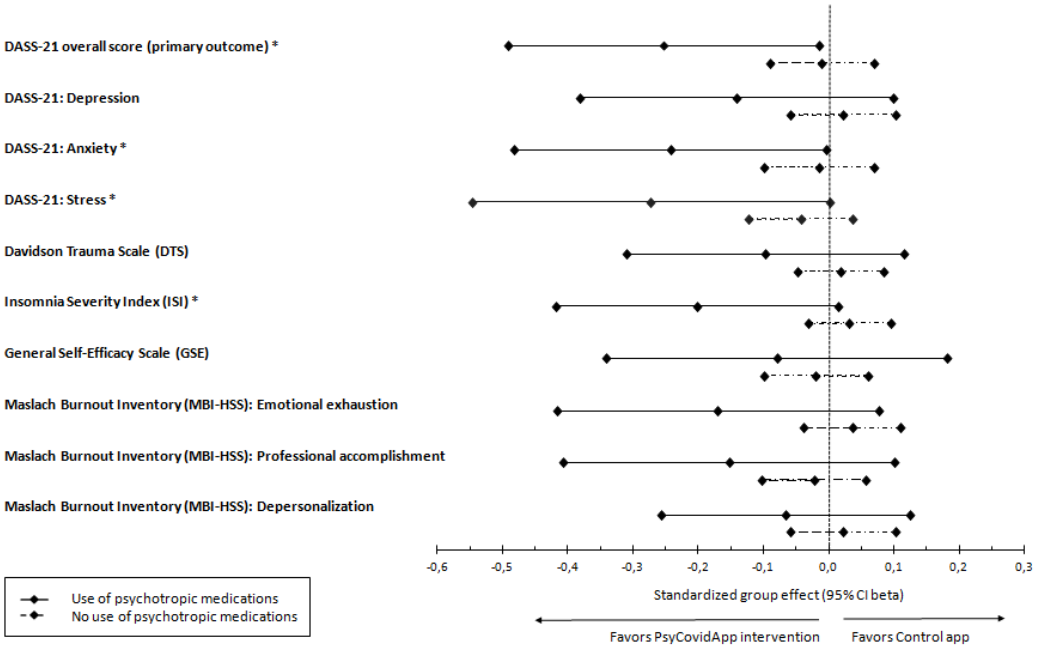
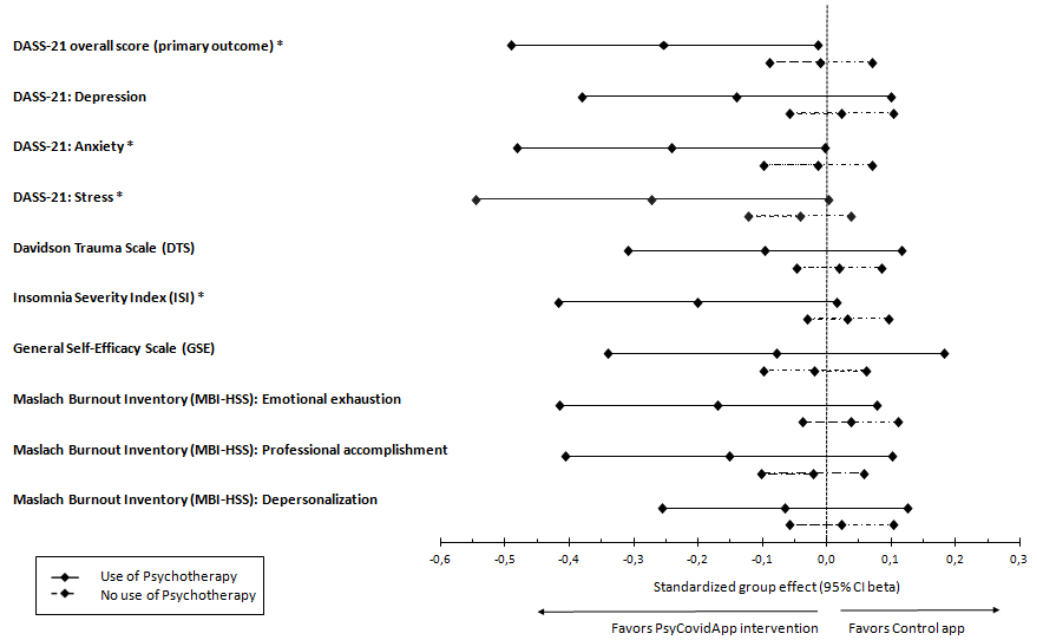


Figure 4. Standardized mean differences for primary and secondary outcomes in health care workers reporting the use of psychotherapy at baseline. Forest plot of standardized group differences between the PsyCovidApp and Control App groups for all outcomes, whereby an effect lower than 0 favored the PsyCovidApp group. Error bars show 95% CIs. DASS-21: Depression, Anxiety, and Stress Scale. *P<.05.



The usability of the PsyCovidApp intervention is described in Table 5. In general, participants perceived that PsyCovidApp was highly usable (mean overall usability score 87.21/100; SD 12.65). After the trial, 208 of the 221 participants in the intervention group (94.1%) asked the research team if they could regain access to the PsyCovidApp intervention.

Table 5. Usability of the PsyCovidApp intervention by the PsyCovidApp group (n=221) measured with the System Usability Scale (the theoretical score range is 0-4 for single items; higher scores are indicative of higher usability).

Item	Mean score (SD)
I would like to use this App frequently	3.00 (1.02)
The App is not unnecessarily complex	3.39 (1.03)
The App is easy to use	3.56 (0.88)
No need for the support of a technical person to use this App	3.73 (0.81)
The various functions in this App are well integrated	3.40 (0.83)
Not too much inconsistency in this App	3.61 (0.81)
Most people would learn to use this App very quickly	3.43 (0.82)
The App is not very cumbersome to use	3.70 (0.70)
Confidence using the App	3.35 (0.95)
No need to learn a lot of things to use the App	3.69 (0.84)
Overall usability score[a]	87.21 (12.65)

[a]Overall System Usability Scale theoretical score range: 0-100 (higher scores are indicative of higher usability).

The sensitivity analysis of all outcomes on a complete case basis (ie, without imputation or adjustment for baseline predictors of missingness) is shown in Table 4. The results of these sensitivity analyses had little effect on the results, with similar findings for the primary outcome and the secondary outcomes when compared with the results of the main analyses.

Discussion

Principal Results

The global health emergency generated by the COVID-19 pandemic is posing an unprecedented challenge to frontline health care workers, who are facing high levels of workload under psychologically difficult situations with scarce resources and support. To our knowledge, this is the first randomized controlled trial to date to assess the efficacy of a mental mHealth intervention for frontline health care workers fighting the health emergency generated by the COVID-19 pandemic. Our analysis showed that at 2 weeks, PsyCovidApp only produced significant improvements in the primary and secondary outcomes of health care workers who were receiving psychotherapy or psychotropic medications.

Implications for Clinical Practice

Although the results of our trial indicate that the PsyCovidApp intervention was not effective in comparison with the control app in the overall health care worker population, we cannot rule out the possibility that the intervention produced beneficial effects that our trial was not able to detect for various reasons, including the choice of an active comparator and the level of use of the intervention. Concerning the choice of an active comparator, most of the outcomes at 2 weeks improved similarly in both the intervention and control group. The improvements in the Control App group could be attributed to the natural progression of the disease in a context of decreasing levels of external stressors (as the impact of the first wave of COVID-19 in Spain was starting to decrease by the time the trial was initiated). However, it is also plausible that the intervention in the control group (which consisted of a similar app but with access limited to general information and contents) also had a positive effect. The impact of the control app may have been enhanced by the Hawthorne effect [40] because the stimulus introduced by the control app could have induced a positive behavioral change due to awareness of being observed. Therefore, we cannot rule out the possibility that the use of a passive comparator (eg, waiting list) in our trial may have resulted in a different outcome for PsyCovidApp.

It is plausible that the intervention did not produce the desired effects because of the short trial duration (ie, too short for the intervention to produce the intended benefits). During the time of the study, health care workers in Spain were overwhelmed with heavy workloads, and it is likely that a large proportion struggled to find time to use PsyCovidApp during the 2-week intervention period. Suboptimal use of mental health apps is indeed a widely acknowledged challenge: user retention rate for smartphone apps in the general population is low, and approximately 25% of users abandon apps after one use [41]. As pointed out by a recent systematic review of apps for depression and anxiety, more than 70% of users stopped using mental health apps after 6 weeks [42]. Unfortunately, we were not able to register the time of use, and we were therefore not able to explore a potential dose-effect relationship.

In any case, as it stands, the trial showed that the PsyCovidApp intervention did not produce significant improvements in the primary and secondary outcomes in the overall population when compared with a control app. It could be interpreted that the PsyCovidApp intervention was not effective in improving mental health outcomes in the short term in this specific population and context. It could be argued that considering all the issues health care workers are required to deal with, providing psychological aid only through a mHealth intervention may not be sufficient to produce significant improvements.

In our subgroup analyses, we observed that the intervention did not produce significant effects among those health care workers using the intervention in absence of additional mental help. This finding is consistent with findings from a recent systematic

review, which showed a lack of effect of mental mHealth interventions when used as a standalone therapy [43]. However, PsyCovidApp was effective in improving the primary outcome and some secondary outcomes when used in conjunction with evidence-based treatments (such as psychotherapy or psychotropic medications). This result supports findings from a recent trial, which showed that a web-based psychoeducation approach was not sufficiently effective in improving depressive symptoms in a general population of workers but was effective for workers who had recently sought help for mental health [11]. It also supports findings from another recent trial, which reported that a web-based psychoeducational intervention produced a significantly greater reduction in depression severity among participants who had undergone psychotherapy before enrolling in the study [12]. According to the ICEMAN criteria (Multimedia Appendix 6), the results of our subgroup are highly credible because they were correctly hypothesized a priori, are supported by prior evidence [11,12], are supported by tests of interaction, are the result of analyzing a small number of effect modifiers, and show consistent effect modification across related outcomes. We hypothesize that the level of motivation to engage with the PsyCovidApp intervention may account for some of the differences in the effects observed among health care workers using and not using psychotherapy or psychotropic medications. Motivation has indeed been identified as a key element in behavior change interventions. Following the transtheoretical model of behavior change [44], health care workers using psychotherapy and psychotropic medications may be more likely in the "action" stage (as they had already made specific overt modifications in acquiring new healthy behaviors) and therefore may have been more motivated to engage with PsyCovidApp, whereas the rest of the health care workers may be more likely in the "precontemplation" or "contemplation" stages (ie, not intending to take action in the foreseeable future) and therefore may have been less motivated to engage with and follow the techniques recommended by PsyCovidApp. Future studies are needed to analyze this hypothesis. However, altogether, our results suggest that although the mHealth interventions may not be effective as a standalone strategy, one possibility to benefit from apps could be to integrate them into a clinical setting and use them in conjunction with other evidence-based treatments. This is a relevant field that is still not well understood and should be further investigated.

The fact that female health care workers were overrepresented in our trial (83% in our trial vs 68% overall in Spain [45]) could have biased our results if the intervention had produced a differential effect by gender. However, as far as we know, there is no evidence in the literature that such a difference exists. In our postprotocol subgroup analyses, we did not find gender differences in any of the outcomes considered (results not shown); therefore, this overrepresentation of female participants is unlikely to have substantially influenced the results of our trial.

PsyCovidApp presented a high usability level, with an overall score of 87.2 points—clearly above the threshold of 68 points used to determine high usability [24]. Usability factors have been widely recognized as key factors to enhance the acceptance of information and communication technologies tools. According to the technology acceptance model, the intention to use a product in the future is strongly correlated with its ease of use [46]. We observed that more than 90% of the participants asked to regain access to the PsyCovidApp intervention after the trial. This finding supports the correlation of ease of use and intention to use, and at the same time, it suggests that the intervention was acceptable and perceived as useful.

In terms of future research needs, it is worth noting that PsyCovidApp was based on a group of software features related to the intervention (eg, learning and in situ use) and communication (eg, prompting) deployed in smartphone interventions that mostly mimic more traditional mobile phone and mHealth solutions. More innovative use of the capabilities of smartphones, such as sensing, alternative delivery paradigms, and advanced analytics, could have produced a more beneficial effect. The possibilities of current smartphone technology have only just been tapped, and further research is needed to explore them fully, as are studies to rigorously analyze the empirical effectiveness of these systems.

A process evaluation is now underway, which will shed light on the mechanisms and contexts in which the intervention did or did not work. In this process evaluation, we will retrospectively investigate the "reach" of interventions (the extent to which the study participants came into contact with the intervention and how they did so). Although a recent meta-analysis of a range of mental health apps concluded that age does not impact treatment effect [19], we cannot rule out the possibility that in our trial, older professionals experienced more problems engaging with PsyCovidApp than younger participants. Therefore, as part of the process evaluation, we will specifically examine the extent to which intervention engagement differed across age groups. We will also use qualitative research methods to gain a deeper understanding of implementation barriers and facilitators and to identify suggestions about how to improve the intervention to maximize its effects on a broader range of health care workers. Once PsyCovidApp becomes publicly available, we will prospectively follow up with new users to identify patterns of use associated with higher intervention benefits. The findings of this process evaluation will inform future developments of the PsyCovidApp intervention.

Limitations

The study has several limitations. First, the 2-week follow-up period may not have been sufficient to detect clinically meaningful differences in mental health. A longer period of time may be needed to produce the desired positive effects. There were two main reasons for this short follow-up period: (1) according to available literature [42], the use of mental health apps substantially decreases after the initial weeks, and (2) the short follow-up allowed us to obtain evidence in a timely manner, which was critical to inform decisions about scaling up the intervention in a time when health care workers in Spain were experiencing remarkably high prevalence rates of stress, anxiety, posttraumatic stress disorder, and insomnia. Second, the mental health of the participants was not evaluated through a diagnostic clinical interview but rather using instruments

indicated for symptomatology assessment rather than for clinical diagnosis. Third, we did not restrict our sample to health care workers with mental health problems at baseline. Including a large proportion of participants with no (or minor) mental health problems in our study may have limited our ability to observe mental health improvements. Fourth, we did not include a waiting list or treatment-as-usual control group; thus, we are unable to determine whether the apparent reduction in mental health symptoms in both groups at 2 weeks represents equal effectiveness of the treatment allocations or the natural progression of the symptoms. Fifth, it was technically unfeasible to monitor use of the intervention, which prevented us to explore a dose-response effect. Finally, our study was remarkably specific to the current pandemic context, which limits the external validity of our results. The trial population was restricted to health care workers who had provided direct health care to patients with COVID-19. The intervention was specifically designed to address the most common mental health problems experienced under these special circumstances, included specific content acknowledging the key challenges health care workers face during the COVID-19 situation, and provided recommendations about how to overcome them. Therefore, and according to best practice guidelines [47], for the investigation of the impact of mobile health interventions in health care workers in a broader, nonpandemic context, the findings of our trial should only be taken into consideration as indirect evidence.

Strengths

The strengths of this study include the pragmatic design, large sample size, and high follow-up rates. Moreover, the trial participants, outcome assessors, and data analysts of the research were blinded to the intervention allocation to reduce biases in the evaluation of the effects of the intervention. A common limitation of previous mHealth trials is that researchers do not have control of the proportion of participants having actual access to their interventions. In our trial, we ensured that all participants successfully downloaded and activated the app before their enrollment in the trial, which is a novel and important strength.

Conclusion

For the first time, the PsyCovidApp trial studied the impact of a cognitive behavioral therapy and mindfulness-based mHealth intervention specifically designed to protect the mental health of health care workers fighting on the front lines of the COVID-19 pandemic. No significant differences were observed between the intervention and control groups at 2 weeks in the primary outcome and in the rest of the outcomes. However, significant improvements were observed among health care workers who were consuming psychotropic medications or receiving psychotherapy in the primary outcome, as well as in posttraumatic stress, insomnia, anxiety, and stress. PsyCovidApp may therefore improve mental health among health care workers who are already using other effective interventions, such as psychotherapy or pharmacological treatments.

Data Sharing

Deidentified data collected for the study, including individual participant data and a data dictionary defining each field in the set, will be made available to others upon request to the corresponding author, following a signed data access agreement.

Acknowledgments

This study was funded by the Balearic Islands Health Research Institute (IdISBa) through the Regional Government of the Balearic Islands, Spain (grant code: COVID-19/06). IRC was supported by a Miguel Servet Fellowship (CP17/00019) from the Instituto de Salud Carlos III (Spanish Ministry of Sciences, Innovation and Universities). MJSR was supported by a FOLIUM - FUTURMed Fellowship from IdISBa, co-founded by ITS-2017 and PO FSE 2014-2020 from the Balearic Islands. We thank our psychologist team (Marian Pérez, Elena Bonet, Toni Riera, Andrea Seguí, Beatriz González, Yasmina Castaño, María Barceló, Catalina Calafat, Miriam Kefauver, Desirée Guillem, Omar Vial, Adela Marmy, Inés Forteza-Rey, Patricia López, Ana Serapio, Catalina Moragues, Alberto Ávila and Ana Martí) for conducting the psychological assessments of the health care workers. We also thank Cristian Sánchez and Daniel Gallego for their support with data management and with technical assistance for health care workers who experienced difficulties in setting up the study Apps. We also thank Massimo de Faveri, Mercedes Martínez, and Alfonso Morillas, from Apploading, Inc, for making the Clinicovery app available for use in this study. Finally, we thank all the health care workers who participated in this study.

Authors' Contributions

IRC, MJSR, MAF, and JGC designed the study. MJSR, AMYJ, MBV, MEGB, MGT, CS, RJ, EG, and JGC developed the contents of the PsyCovidApp intervention under the coordination of MJSR. MAF transferred the intervention contents to the Clinicovery app. MJSR and the psychologist team (see *Acknowledgments*) collected the data. MAF conducted data management. RJ, EG, AYMJ, and AL analyzed the data. All authors interpreted the data. IRC wrote the report. All authors critically revised and approved the final version of the manuscript. IRC and MJSR were the principal co-investigators.

Conflicts of Interest

MJSR, AMYJ, MBV, MEGB, MGT, CS, RJ, EG, and JGC developed the contents of the intervention. All other authors have no conflicts to declare.

Multimedia Appendix 1
Content of the PsyCovidApp intervention.
[PDF File (Adobe PDF File), 777 KB - mhealth_v9i5e27039_app1.pdf]

Multimedia Appendix 2
Content of the control app.
[PDF File (Adobe PDF File), 110 KB - mhealth_v9i5e27039_app2.pdf]

Multimedia Appendix 3
Subgroup analyses: comparison of outcome measures between the PsyCovidApp and the Control App groups at 2 weeks in prespecified subgroups (high DASS-21 baseline scores; users of psychotherapy; consumers of psychotropic medications).
[PDF File (Adobe PDF File), 202 KB - mhealth_v9i5e27039_app3.pdf]

Multimedia Appendix 4
Number of health care workers recruited by Spanish region.
[PDF File (Adobe PDF File), 96 KB - mhealth_v9i5e27039_app4.pdf]

Multimedia Appendix 5
Trial results reported in terms of adjusted mean between-group differences (overall sample).
[PDF File (Adobe PDF File), 133 KB - mhealth_v9i5e27039_app5.pdf]

Multimedia Appendix 6
Instrument to Assess the Credibility of Effect Modification Analyses (ICEMAN) used in the study.
[PDF File (Adobe PDF File), 160 KB - mhealth_v9i5e27039_app6.pdf]

Multimedia Appendix 7
CONSORT-EHEALTH checklist (v 1.6.1).
[PDF File (Adobe PDF File), 1102 KB - mhealth_v9i5e27039_app7.pdf]

References

1. COVID-19 Weekly Epidemiological Update. World Health Organization. 2020 Nov 15. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update---17-november-2020 [accessed 2021-05-11]
2. Kang L, Li Y, Hu S, Chen M, Yang C, Yang BX, et al. The mental health of medical workers in Wuhan, China dealing with the 2019 novel coronavirus. Lancet Psychiatry 2020 Mar;7(3):e14 [FREE Full text] [doi: 10.1016/S2215-0366(20)30047-X] [Medline: 32035030]
3. Serrano-Ripoll MJ, Meneses-Echavez JF, Ricci-Cabello I, Fraile-Navarro D, Fiol-deRoque MA, Pastor-Moreno G, et al. Impact of viral epidemic outbreaks on mental health of healthcare workers: a rapid systematic review and meta-analysis. J Affect Disord 2020 Dec 01;277:347-357 [FREE Full text] [doi: 10.1016/j.jad.2020.08.034] [Medline: 32861835]
4. Kisely S, Warren N, McMahon L, Dalais C, Henry I, Siskind D. Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis. BMJ 2020 May 05;369:m1642 [FREE Full text] [doi: 10.1136/bmj.m1642] [Medline: 32371466]
5. Bandyopadhyay S, Baticulon RE, Kadhum M, Alser M, Ojuka DK, Badereddin Y, et al. Infection and mortality of healthcare workers worldwide from COVID-19: a systematic review. BMJ Glob Health 2020 Dec;5(12) [FREE Full text] [doi: 10.1136/bmjgh-2020-003097] [Medline: 33277297]
6. Luceño-Moreno L, Talavera-Velasco B, García-Albuerne Y, Martín-García J. Symptoms of posttraumatic stress, anxiety, depression, levels of resilience and burnout in Spanish health personnel during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Jul 30;17(15) [FREE Full text] [doi: 10.3390/ijerph17155514] [Medline: 32751624]
7. Alexopoulos AR, Hudson JG, Otenigbagbe O. The use of digital applications and COVID-19. Community Ment Health J 2020 Oct;56(7):1202-1203 [FREE Full text] [doi: 10.1007/s10597-020-00689-2] [Medline: 32734311]
8. Davidson SK, Schattner PL. Doctors' health-seeking behaviour: a questionnaire survey. Med J Aust 2003 Sep 15;179(6):302-305. [doi: 10.5694/j.1326-5377.2003.tb05552.x] [Medline: 12964913]
9. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
10. Andersson E, Steneby S, Karlsson K, Ljótsson B, Hedman E, Enander J, et al. Long-term efficacy of Internet-based cognitive behavior therapy for obsessive-compulsive disorder with or without booster: a randomized controlled trial. Psychol Med 2014 Oct;44(13):2877-2887. [doi: 10.1017/S0033291714000543] [Medline: 25066102]

11. Imamura K, Kawakami N, Tsuno K, Tsuchiya M, Shimada K, Namba K. Effects of web-based stress and depression literacy intervention on improving symptoms and knowledge of depression among workers: A randomized controlled trial. J Affect Disord 2016 Oct;203:30-37. [doi: 10.1016/j.jad.2016.05.045] [Medline: 27280960]
12. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? Efficacy of Internet-based guided self-help compared to online psychoeducation for major depressive disorder. J Affect Disord 2019 Mar 01;246:695-705. [doi: 10.1016/j.jad.2018.12.065] [Medline: 30611913]
13. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
14. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR mHealth uHealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
15. Steinkamp JM, Goldblatt N, Borodovsky JT, LaVertu A, Kronish IM, Marsch LA, et al. Technological interventions for medication adherence in adult mental health and substance use disorders: a systematic review. JMIR Ment Health 2019 Mar 12;6(3):e12493 [FREE Full text] [doi: 10.2196/12493] [Medline: 30860493]
16. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR mHealth uHealth 2017 Jul 20;5(7):e94 [FREE Full text] [doi: 10.2196/mhealth.7088] [Medline: 28729235]
17. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR mHealth uHealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
18. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
19. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
20. Drissi N, Ouhbi S, Marques G, de la Torre Díez I, Ghogho M, Janati Idrissi MA. A systematic literature review on e-mental health solutions to assist health care workers during COVID-19. Telemed J E Health 2020 Sep 22. [doi: 10.1089/tmj.2020.0287] [Medline: 32970532]
21. Pollock A, Campbell P, Cheyne J, Cowie J, Davis B, McCallum J, et al. Interventions to support the resilience and mental health of frontline health and social care professionals during and after a disease outbreak, epidemic or pandemic: a mixed methods systematic review. Cochrane Database Syst Rev 2020 Nov 05;11:CD013779. [doi: 10.1002/14651858.CD013779] [Medline: 33150970]
22. Roland M, Torgerson DJ. What are pragmatic trials? BMJ 1998 Jan 24;316(7127):285-285 [FREE Full text] [doi: 10.1136/bmj.316.7127.285] [Medline: 9472515]
23. Serrano-Ripoll MJ, Ricci-Cabello I, Jiménez R, Zamanillo-Campos R, Yañez-Juan AM, Bennasar-Veny M, et al. Effect of a mobile-based intervention on mental health in frontline healthcare workers against COVID-19: Protocol for a randomized controlled trial. J Adv Nurs 2021 Mar 06. [doi: 10.1111/jan.14813] [Medline: 33675247]
24. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
25. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assess 1998 Jun;10(2):176-181. [doi: 10.1037/1040-3590.10.2.176]
26. Antúnez Z, Vinet E. Escalas de Depresión, Ansiedad y Estrés (DASS - 21): Validación de la Versión abreviada en Estudiantes Universitarios Chilenos. Ter Psicol 2012 Dec;30(3):49-55. [doi: 10.4067/S0718-48082012000300005]
27. Bobes J, González M, Sáiz P, Bascarán M, Bousoño M. Instrumentos básicos para la práctica de la psiquiatría clínica. Barcelona, Spain: Psiquiatría Editores, SL; 2004.
28. Bobes J, Calcedo-Barba A, García M, François M, Rico-Villademoros F, González MP, Grupo Espanol de Trabajo para el Estudio del Trastornostres Postraumático. Evaluation of the psychometric properties of the Spanish version of 5 questionnaires for the evaluation of post-traumatic stress syndrome. Article in Spanish. Actas Esp Psiquiatr 2000;28(4):207-218. [Medline: 11116791]
29. Gil-Monte PR. Factorial validity of the Maslach Burnout Inventory (MBI-HSS) among Spanish professionals. Rev Saude Publica 2005 Feb;39(1):1-8 [FREE Full text] [doi: 10.1590/s0034-89102005000100001] [Medline: 15654454]
30. Olivares-Faúndez V, Mena-Miranda L, Macía-Sepúlveda F, Jélvez-Wilke C. Validez factorial del Maslach Burnout Inventory Human Services (MBI-HSS) en profesionales chilenos. Univ Psychol 2014 May 01;13(1). [doi: 10.11144/javeriana.upsy13-1.vfmb]
31. Bastien CH, Vallières A, Morin CM. 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]
32. Sierra JC, Guillén-Serrano V, Santos-Iglesias P. Insomnia Severity Index: some indicators about its reliability and validity on an older adults sample. Article in Spanish. Rev Neurol 2008;47(11):566-570 [FREE Full text] [Medline: 19048535]

33. Herrero R, Espinoza M, Molinari G, Etchemendy E, Garcia-Palacios A, Botella C, et al. Psychometric properties of the General Self Efficacy-12 Scale in Spanish: general and clinical population samples. Compr Psychiatry 2014 Oct;55(7):1738-1743. [doi: 10.1016/j.comppsych.2014.05.015] [Medline: 24973225]
34. Soler Sánchez MI, Meseguer de Pedro M, García Izquierdo M. Propiedades psicométricas de la versión española de la escala de resiliencia de 10 ítems de Connor-Davidson (CD-RISC 10) en una muestra multiocupacional. Revista Latinoamericana de Psicología 2016 Sep;48(3):159-166. [doi: 10.1016/j.rlp.2015.09.002]
35. Orellana C, Orellana L. Propiedades psicométricas de la Escala abreviada de Depresión, Ansiedad y Estrés (DASS-21) en adultos salvadoreños. Teoría y Praxis 2021;19(38):38-54 [FREE Full text]
36. Bados A, Solanas A, Andrés R. Psychometric properties of the Spanish version of Depression, Anxiety and Stress Scales (DASS). Psicothema 2005;17(4):679-683 [FREE Full text]
37. Gao M. Multi-Cultural Usability Assessment with System Usability Scale. Thesis. Rice University. 2019. URL: https://hdl.handle.net/1911/106168 [accessed 2021-05-11]
38. Royston P, White I. Multiple imputation by chained equations (MICE): implementation in Stata. J Stat Soft 2011;45(4). [doi: 10.18637/jss.v045.i04]
39. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the instrument to assess the credibility of effect modification analyses (ICEMAN) in randomized controlled trials and meta-analyses. CMAJ 2020 Aug 10;192(32):E901-E906 [FREE Full text] [doi: 10.1503/cmaj.200077] [Medline: 32778601]
40. Wickström G, Bendix T. The "Hawthorne effect"—what did the original Hawthorne studies actually show? Scand J Work Environ Health 2000 Aug;26(4):363-367 [FREE Full text] [Medline: 10994804]
41. Rodde T. 25% of users abandon an app after one use. Upland. 2019. URL: http://info.localytics.com/blog/25-of-users-abandon-apps-after-one-use [accessed 2021-05-11]
42. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
43. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
44. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychol Psychother 1982;19(3):276-288. [doi: 10.1037/h0088437]
45. Profesionales sanitarios colegiados. Instituto Nacional de Estadística. 2019. URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176781&menu=ultiDatos&idp=1254735573175 [accessed 2021-05-11]
46. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
47. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

DASS-21: Depression, Anxiety, and Stress Scale
DTS: Davidson Trauma Scale
GSE: General Self-Efficacy Scale
IdISBa: Balearic Islands Health Research Institute
ISI: Insomnia Severity Index
MBI-HSS: Maslach Burnout Inventory - Human Services Survey
mHealth: mobile health
SUS: System Usability Scale

Edited by L Buis; submitted 11.01.21; peer-reviewed by N Kawakami, T Alessa; comments to author 16.02.21; revised version received 08.03.21; accepted 31.03.21; published 18.05.21.

Please cite as:
Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, Zamanillo-Campos R, Yáñez-Juan AM, Bennasar-Veny M, Leiva A, Gervilla E, García-Buades ME, García-Toro M, Alonso-Coello P, Pastor-Moreno G, Ruiz-Pérez I, Sitges C, García-Campayo J, Llobera-Cánaves J, Ricci-Cabello I
A Mobile Phone–Based Intervention to Reduce Mental Health Problems in Health Care Workers During the COVID-19 Pandemic (PsyCovidApp): Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(5):e27039
URL: https://mhealth.jmir.org/2021/5/e27039
doi:10.2196/27039
PMID:33909587

Original Paper

Feasibility of a Waistband-Type Wireless Wearable Electrocardiogram Monitoring System Based on a Textile Electrode: Development and Usability Study

Danbi Gwon[1], BEng; Hakyung Cho[2*], PhD; Hangsik Shin[1*], PhD

[1]Department of Biomedical Engineering, Chonnam National University, Yeosu, Republic of Korea

[2]BLACKYAK Co Ltd, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hangsik Shin, PhD
Department of Biomedical Engineering
Chonnam National University
50 Daehak-ro
Yeosu, 59626
Republic of Korea
Phone: 82 61 659 7362
Email: hangsik.shin@jnu.ac.kr

Abstract

Background: Electrocardiogram (ECG) monitoring in daily life is essential for effective management of cardiovascular disease, a leading cause of death. Wearable ECG measurement systems in the form of clothing have been proposed to replace Holter monitors used for clinical ECG monitoring; however, they have limitations in daily use because they compress the upper body and, in doing so, cause discomfort during wear.

Objective: The purpose of this study was to develop a wireless wearable ECG monitoring system that includes a textile ECG electrode that can be applied to the lining of pants and can be used in the same way that existing lower clothing is worn, without compression to the upper body.

Methods: A textile electrode with stretchable characteristics was fabricated by knitting a conductive yarn together with polyester-polyurethane fiber, which was then coated with silver compound; an ECG electrode was developed by placing it on an elastic band in a modified limb lead configuration. In addition, a system with analog-to-digital conversion, wireless communication, and a smartphone app was developed, allowing users to be able to check and store their own ECGs in real time. A signal processing algorithm was also developed to remove noise from the obtained signal and to calculate the heart rate. To evaluate the ECG and heart rate measurement performance of the developed module, a comparative evaluation with a commercial device was performed. ECGs were measured for 5 minutes each in standing, sitting, and lying positions; the mean absolute percentage errors of heart rates measured with both systems were then compared.

Results: The system was developed in the form of a belt buckle with a size of 53 × 45 × 12 mm (width × height × depth) and a weight of 23 g. In a qualitative evaluation, it was confirmed that the P-QRS-T waveform was clearly observed in ECGs obtained with the wearable system. From the results of the heart rate estimation, the developed system could track changes in heart rate as calculated by a commercial ECG measuring device; in addition, the mean absolute percentage errors of heart rates were 1.80%, 2.84%, and 2.48% in the standing, sitting, and lying positions, respectively.

Conclusions: The developed system was able to effectively measure ECG and calculate heart rate simply through being worn as existing clothing without upper body pressure. It is anticipated that general usability can be secured through further evaluation under more diverse conditions.

(JMIR Mhealth Uhealth 2021;9(5):e26469) doi:10.2196/26469

KEYWORDS

electrocardiogram; telehealth; telemetry; telemonitoring; textile electrode; wearable system; smartphone; mobile phone

Introduction

According to the World Health Organization, cardiovascular disease is the leading cause of death worldwide, accounting for 31% of all deaths [1]. Arrhythmia is a problem with the rate or rhythm of the heartbeat and is the most common and dangerous of the cardiovascular diseases. It has been reported that 15% of adults with congenital cardiovascular disease have arrhythmias, the condition occurs in 7% of patients in their 20s and 38% of patients in their 50s, and the incidence continues to increase with age [2]. Arrhythmia can cause sudden death in severe cases, and it is known that 25% of deaths from heart disease are caused by cardiac arrhythmias [3]. They are difficult to completely cure and need to be permanently managed throughout life [4]. One of the most difficult aspects of the diagnosis and treatment of arrhythmia is that it does not always persist, can occur at any point, and can disappear intermittently. Therefore, some patients are required to record their electrocardiogram (ECG) continuously for extended periods, and for this, 24-hour ECG monitoring with a Holter monitor is used clinically to accurately track and diagnose the occurrence of arrhythmias. Holter monitoring is a technology that records an ECG through standard limb leads using a small portable measuring device. A silver-silver chloride electrode is usually attached to the chest, and then an ECG is recorded by wired connection between an electrode and a necklace-type or a waist-worn–type signal acquisition device. Holter monitoring has been used for many years for precise diagnosis of arrhythmia but has the following limitations. First, self-use of Holter monitors are limited since the user's understanding of the ECG lead is required when attaching electrodes. In addition, since recording is performed by a bulky device that needs to be placed on the chest or waist by a wired connection to the electrode on the chest, it may cause user discomfort. Furthermore, there are burdens associated with aesthetics (ie, some find them unattractive), or privacy related to an individual's disease state may be compromised. Lastly, since Holter monitoring devices are used mainly for clinical diagnostic purposes, they are not suitable for users who want to monitor their ECG during normal daily life for the purposes of prevention, early diagnosis, and prognosis of cardiovascular diseases. Therefore, there is a growing need for novel ECG measurement technologies to expand the universe of users and to improve the discomfort associated with Holter devices. Since it is critical that the user always carries a device in order to continuously measure an ECG, wearable technology has been suggested as the most promising always-on approach. The clothing-type ECG system, which is one of several representative wearable technologies, embeds an ECG electrode and the attendant system into clothing. In general, these systems attach a conductive textile electrode made through a conductive fiber to a compressive top and connects it with a separate measuring device to obtain the physiological signal [5-16].

As an example of a clothing-type ECG system, Takahashi and Suzuki proposed a system for measuring ECGs that involved attaching conductive textile electrodes to both sleeve cuffs [17]. In another example, a conductive fiber was embroidered at the position of an ECG electrode on a top, and the signal was then recorded by connecting embroidered conductive yarn to a measurement system [18]. In this approach, a conductive fiber was also embroidered between the textile electrode and the system and used as an electrical wire. In addition, there is a method for measuring the user's ECG in a noncontact manner by attaching a capacitive ECG electrode to the outside of clothing [19]. As described, the clothing-type ECG systems do not need a sensor attachment to the body but are simply integrated with the clothing to be worn in the same manner, thus improving user convenience. However, in terms of practical use, these systems have the following limitations. First, a textile electrode used in a clothing-type system has a lower impedance and lower contact stability than a medical grade electrode, and this may degrade the quality of the signal to be measured. In addition, since the electrode is generally located in or around a specific part of the upper body that experiences a lot of movement (eg, heart and upper limbs), it is vulnerable to motion artifacts.

In order to solve these problems, a method of improving the contact force of the sensor by using clothing in a form that is highly elastic and compressive to the body has been proposed, but the limitation with this approach is the significant discomfort caused by sustained compression of the user's body. Therefore, a compelling need exists to develop a new ECG measurement technology to solve some of the problems of the existing clothing-type wearable systems.

This study aimed to develop a wireless wearable ECG monitoring system that included a textile ECG electrode that can be applied to the lining of pants and can be used in the same way as existing clothing wearing styles without upper body compression. Since the electrode is in close contact with the skin through the usual manner of wearing lower garments—tightening at the waist to make the clothes ride at the waist or pelvis—the contact force between the electrode and the skin can be improved without pressure to the upper torso with the conventional clothing-type ECG method. In addition, there is the advantage of being able to perform this monitoring simply through the normal daily activity of dressing without additional manipulations to attach electrodes. The measurement system we have developed is in the form of a belt buckle, and users can check their ECG as needed through a smartphone app.

Methods

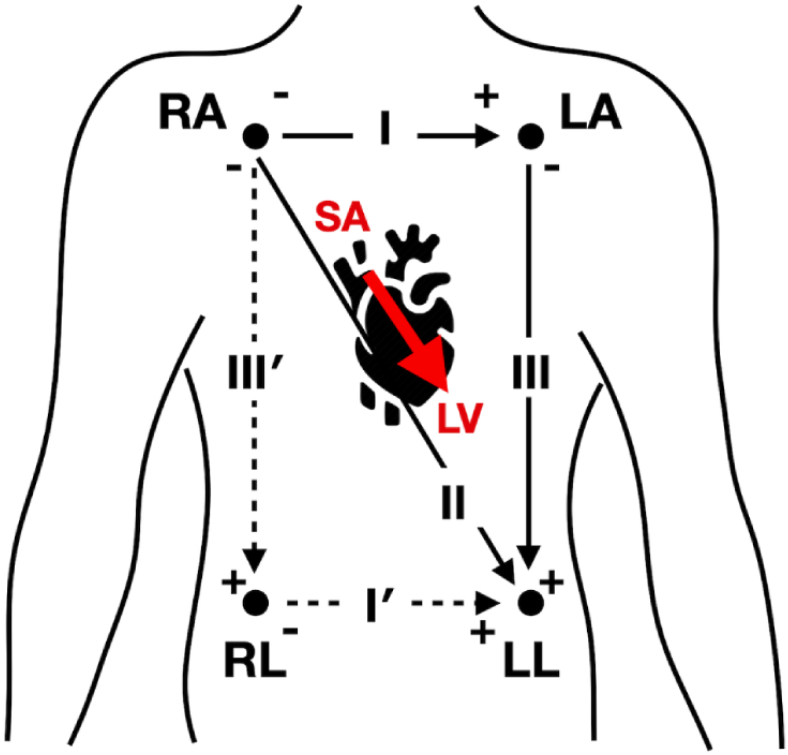
Modified Limb Leads

An ECG measures the electrical activity of the heart on the surface of the human body and refers it to an electrical vector generated by the heart. The most representative method for performing an ECG measurement is by use of the standard limb leads developed by Einthoven. In the standard limb leads system, a total of four electrodes are attached to the arms and legs by crossing the heart, and an ECG is induced from the potential difference between the electrodes. The ECG can be measured three different ways based on the reference position for obtaining the potential difference. The ECG is recorded from the potential difference between the right and left arm for lead I, between the right and left leg for lead II, and between the left arm and left leg for lead III. The rationale behind this lead system is that bioelectricity in the heart is propagated from the right upper

limb to the left lower limb. Intracardiac electrical conduction begins at the sinoatrial (SA) node in the upper wall of the right atrium, passes through the atrioventricular (AV) node located between the atrium and the ventricle, passes through this bundle, and is delivered to Purkinje fibers, a type of fine nerve fiber distributed in the ventricle. Therefore, the directionality of electrical conduction of the heart occurs from the head to the foot; in addition, since the ventricular wall of the left ventricle responsible for systemic circulation is thicker and the distribution of Purkinje fibers is higher, the electrical conduction vector is skewed to the left rather than the right. As a result, electrical conduction of the entire heart occurs from the SA node to the left ventricle, which corresponds to the direction from the right arm to the left leg. Figure 1 shows the direction of electrical conduction and the ECG lead system. The solid diagonal line in Figure 1 shows the standard limb lead of Einthoven's triangle. Lead II is the most similar to the electrical conduction direction of the heart, and it is here that the ECG with the largest amplitude is observed. Lead II can be expressed as a vector sum of horizontal and vertical directions, and these can be defined as leads I and III, respectively.

The standard limb leads should have at least one electrode attached to an upper limb for measurement; thus, conventional clothing-type wearable devices have been mainly developed in a form that has electrodes attached to the top. However, attaching the electrode to the upper body may cause discomfort due to upper limb compression. To solve this problem, we decided to move all electrodes to the bottom, and for this we used a form of modified limb leads. In Figure 1, the vector corresponding to lead II can also be expressed as the sum of vectors I' and III'; in this case, the potential difference between both lower limbs can be measured through lead I'. This lead system does not pressurize or attach an additional electrode to the upper body, so it does not restrict activity and provides the same user experience as wearing items of clothing like pants. As such, it is more convenient than existing wearable ECG systems that attach an electrode on the upper limb. This type of lead system has been previously defined, and the leads are termed "modified limb leads" in the studies of John and Sivaraman [20] and Sivaraman et al [21]. In their studies, they report that the modified limb lead is more useful for observing ECG amplitude or frontal plane axis shift than standard limb leads.

Figure 1. Modified limb leads (I, II, and III). LA: left arm; LL: left leg; LV: left ventricle; RA: right arm; RL: right leg; SA: sinoatrial.



Waistband-Type ECG Electrodes

Fabrication Process for the Textile Electrode

To develop the textile electrode, conductive fibers were made by covering a 70-denier polyester core fiber with two strands of 30-μm diameter silver fibers. In order to obtain elasticity and a close fit of the electrode, a textile electrode knitted with conductive fiber possessing a striped structure was prepared from 18% polyurethane, 82% polyester, and the conductive fiber. For better contact with the skin and more accurate measurement, the knitted textile electrode was coated with a silver compound. The silver compound was selectively printed on the conductive fibers with a silk screen then cured at 100 ˚C for 30 minutes. Silver coating was limited to the area of the knitted fabric containing the conductive fibers, based on considerations of elasticity, conductivity, close contact, possible irritation, and sewing productivity (Figure 2, A). Panels B and C in Figure 2 show the structure of the knitted textile and the knitted textile electrode coated with silver compound,

respectively. The silver compound used in this study is a conductive resin, XX-ECA05 (CEMEDINE Co Ltd), which is composed of elastomeric resin from acrylic resin, diethyl ether, and ethylene glycol derivatives as well as silver and its water-soluble compounds. The composition and properties of XX-ECA05 are shown in Table 1.

Figure 2. Fabrication process and configuration of the knitted textile electrode for electrocardiogram measurement. A. Fabrication process of the knitted textile electrode. B. Structure of the knitted textile. C. Knitted textile electrode coated with silver compound.

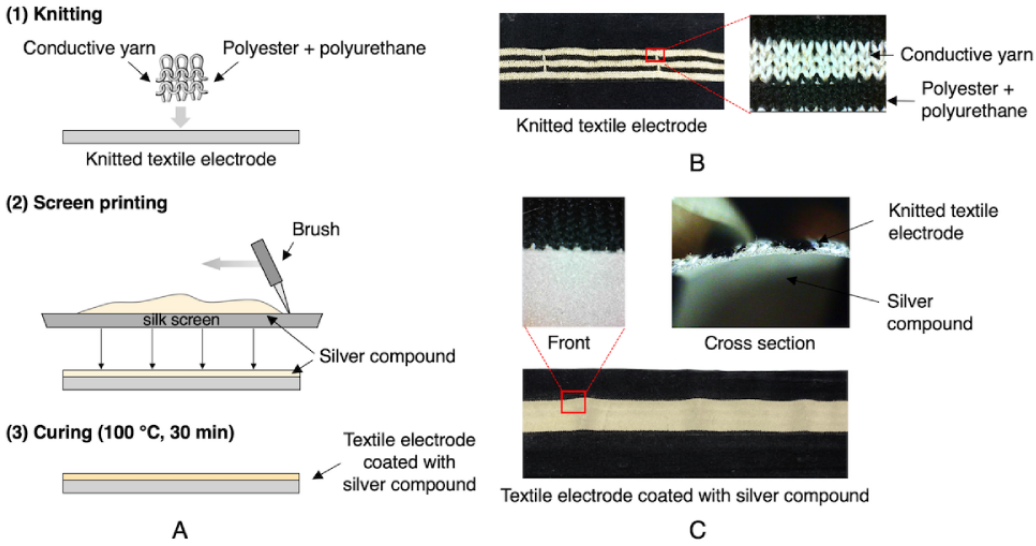


Table 1. Ingredients and material properties of the conductive resin XX-ECA05 (CEMEDINE Co Ltd).

Ingredient name or property	Content, %	Description
Ingredient		
Acrylic resin	10-20	N/A[a]
Diethyl ether	1-10	N/A
Ethylene glycol derivatives	1-5	N/A
Silver and its water-soluble compounds	69	N/A
Material property		
Usage	N/A	Printable, stretchable wiring
Base resin	N/A	Elastomer
Viscosity	N/A	60.0 Pa·s
Curing condition	N/A	100 ˚C × 30 min
Working temperature	N/A	–20 ˚C to 80 ˚C
Volume resistivity	N/A	0.45 mΩ·cm
Elongation	N/A	100%

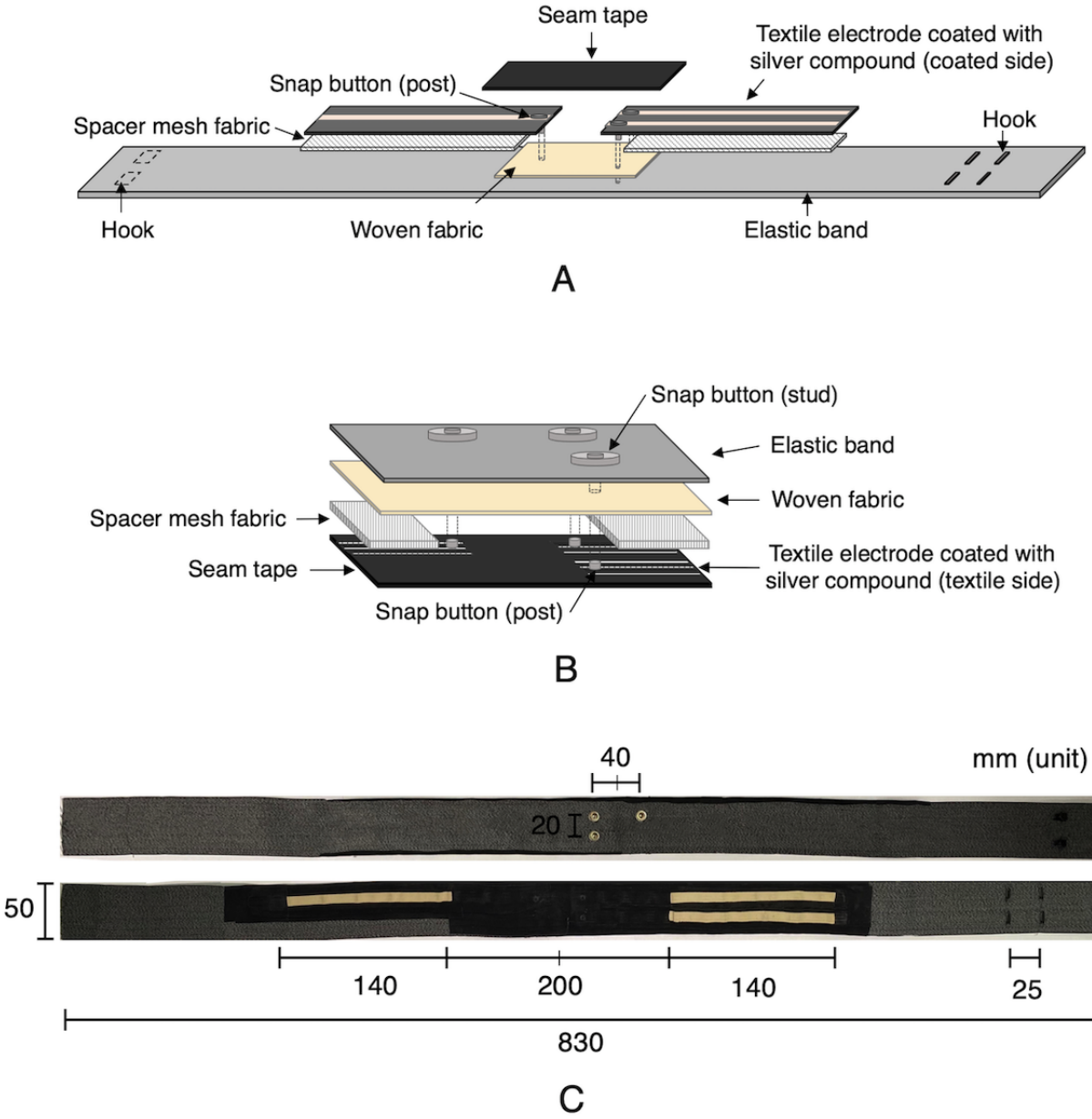
[a]N/A: not applicable; content was only determined for ingredients and descriptions were only provided for properties.

Electrode Configuration on Waistband

In order to fit various sizes and shapes of potential users, a waistband-type ECG lead system was proposed based on an elastic band and the knitted textile electrode. The knitted textile electrode was positioned on the elastic band according to the modified standard limb lead I. An elastic mesh fabric spacer was placed between the elastic band and the knitted textile electrode, thus allowing better contact with the skin [7]. The electrode and the signal acquisition device were connected using a nickel snap button. To ensure that the metallic surface of the snap button would not directly touch the skin, the back surface was covered with seam tape. To prevent deformation of the waistband by the snap button where it connects to the measuring device, a nonelastic woven fabric was placed between the knitted textile electrode and the elastic band (Figure 3, A and B). The length of the elastic band was 830 mm; this was based on the 2015 Size Korea survey, which revealed the average waist size of 25-to-29-year-old subjects was 830.1 (SD 88.5) mm [22]. A two-step size adjustment was made possible by attaching size-adjusting hooks 25 mm apart. The size of the developed waistband was 830 mm long and 50 mm wide; the knitted textile electrodes that were cut to a length of 220 mm were symmetrically placed 20 mm apart from the horizontal center line in the front of the clothing construction to connect with devices. The snap button was positioned 40 mm away

horizontally and 20 mm away vertically from the center front line. Thus, each electrode was exposed starting 100 mm from the center line, and an electrode length of 140 mm was in contact with the skin (Figure 3, C).

Figure 3. Waistband-type electrocardiogram (ECG) lead system using knitted textile electrodes. A. Configuration diagram of waistband for ECG measurement. B. Configuration diagram of snap button for connection between knitted textile electrode and ECG measurement device. C. Appearance of manufactured waistband-type ECG lead system.



Wireless ECG Measurement System

The wearable ECG system was connected to a waistband-type electrode through a snap button, measured the ECG, and transmitted it wirelessly to an Android smartphone. The app displayed the received ECG as a real-time graph and stored it as a file. The data acquisition system consisted of an ADS1191 (Texas Instruments Inc), which is a commercial, analog front end for ECG measurement; an Arduino Pro Mini (SparkFun Electronics), which is an Atmega328-based microcontroller unit (MCU); and an HC-06 Bluetooth communication module (Guangzhou HC Information Technology Co Ltd). The ADS1191 is a 1-channel analog front end that amplifies the ECG and converts the analog signal to digital. The common mode rejection ratio of the ADS1191 is –95 dB and it supports a sampling frequency of 125 to 8000 samples per second with 16-bit analog-to-digital conversion resolution. The Arduino Pro Mini is a microcontroller that operates at up to 16 MHz and supports 14 digital inputs and four analog inputs. The size of the Arduino Pro Mini, at 18 × 33 mm, is suitable for developing small systems. The HC-06 Bluetooth module supports Bluetooth version 2.0 serial port profile and transmits ECG data received from the MCU to the smartphone. Serial peripheral interface communication between the ADS1191 and the Arduino Pro Mini, as well as universal asynchronous receiver/transmitter communication between them, are used for data communication. Table 2 summarizes the system specifications. The buckle-type system cover was designed using 123D Design, version 2.2.14

(Autodesk), a freeware computer-aided design software application, and was printed using the Replicator2 3D printer (MakerBot). The polylactic acid filament, a transparent plastic material, was used to make the cover. The Android app for ECG self-monitoring was developed to run on the SM-N920K smartphone (Samsung Electronics, Inc) based on Android Software Development Kit 24 using Android Studio, version 3.5.3. After receiving data from the digital system's Bluetooth module, the app would display the ECG waveform in real time and store the received ECG data as a comma-separated value format file to the device storage space. The MPAndroidChart library, a graph generation library, was used to display real-time ECG waveforms [23].

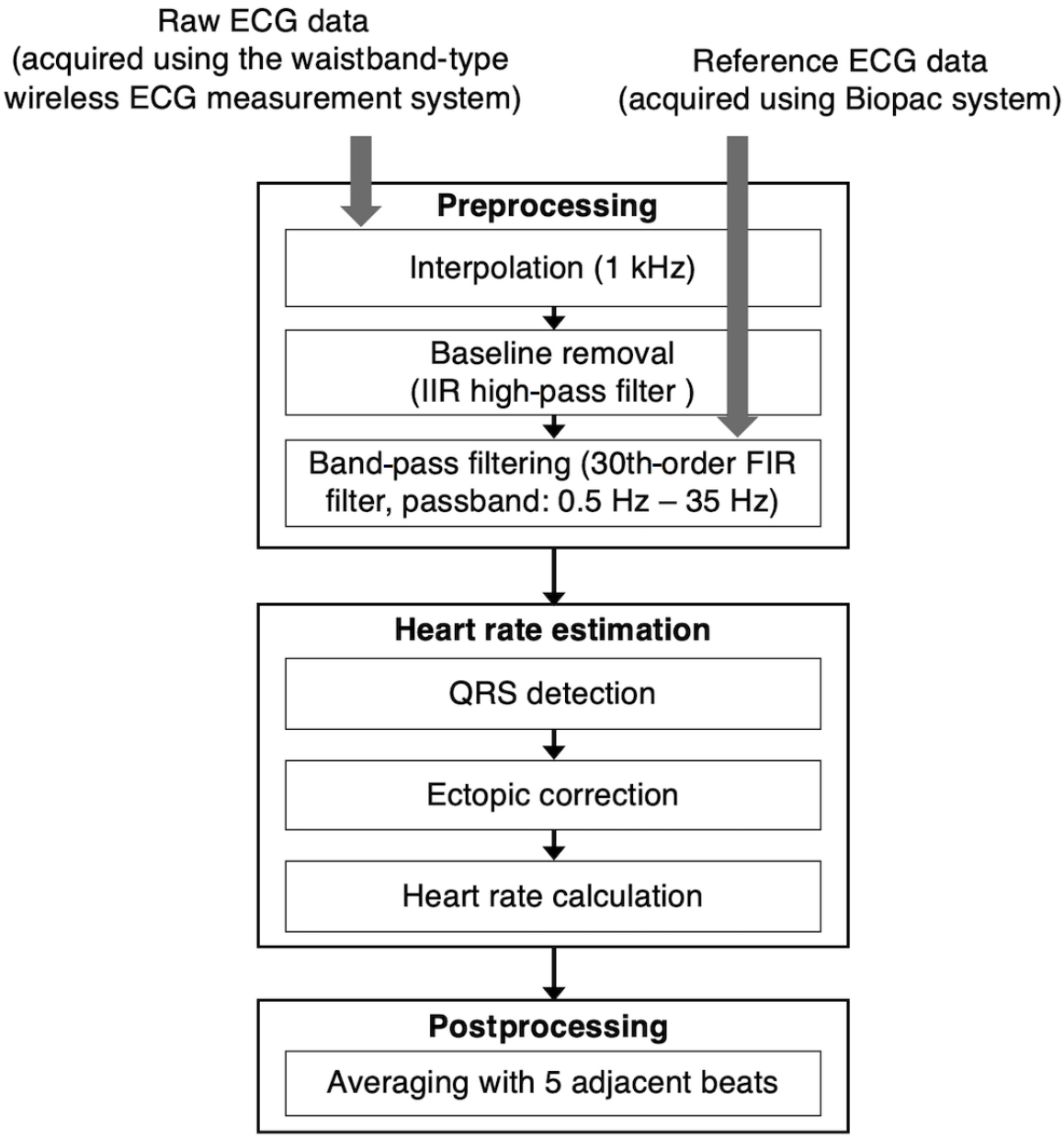
Table 2. Specifications of the developed system.

Module and specification	Value
Analog front end (ADS1191)	
Analog-to-digital conversion resolution	16 bits
Sampling rate	125 samples per second
Microcontroller (Arduino Pro Mini)	
Clock speed	16 MHz
Operating voltage	3.3 V
Bluetooth module (HC-06)	
Bluetooth version	V2.0
Operating voltage	3.2-6 V
Operating frequency	2.4 GHz
Total system, including cover and battery	
Size (width × height × depth)	53 × 45 × 12 mm
Weight	23 g

Signal Processing Algorithm for Heart Rate Estimation

The waistband-type ECG system does not use adhesive foam for attachment to the human body, and the electrode-skin contact state may vary drastically by movement, so compared to existing ECG systems, it is quite vulnerable to motion artifacts, including noise from respiratory movement. In addition, data may be lost due to communication delay or packet loss during wireless data transmission. Therefore, in order to accurately extract the QRS complex and estimate heart rate, a signal processing procedure for noise removal and signal restoration is required. Figure 4 shows the algorithm for extracting heart rate from the waistband-measured ECG. First, since a transmission delay or loss may occur in the process of receiving a signal, the obtained signal was interpolated at 1 kHz to compensate for uneven spacing of the signal. Then, an infinite impulse response high-pass filter was applied to reduce baseline fluctuation. Thereafter, a 30^{th}-order finite impulse response filter having a 0.5 Hz to 35 Hz passband known as a general ECG frequency band was applied. After the preprocessing stage, QRS complexes were detected using the Pan-Tompkins algorithm [24], then the RR interval (RRI) was calculated based on the detected QRS. At this time, a QRS detection error may occur due to an incorrect QRS detection or transmission error, and this may cause a phenomenon in which the RRI has an out-of-range value that is abnormal. This can lead to a fatal error in calculating the heart rate. Therefore, after calculation of the RRI, abnormal range value correction is required. To this end, the RRI was reconstructed using the algorithm proposed by Morelli et al [25]. In this algorithm, the number of lost beats is estimated from the abnormal value; from this, the average RRI value and the lost beats are estimated and restored. After correction, the heart rate is calculated by taking its reciprocal. When the heart rate is calculated for each RRI, it may appear that it changes very rapidly due to a subtle error in the QRS position. For example, when calculating the heart rate with only a one-beat interval, if the QRS position is measured outside of 50 milliseconds due to a detection error, an error of about 5 beats per minute (bpm) may occur based on a heart rate of 72 bpm. Therefore, in order to reduce the influence of these minute fluctuations, it is common to obtain several adjacent beats and provide an average beat rate. In this study, the average bpm was calculated from the average of 5 beats, and for this, a 5-point moving average filter (MAF) was applied.

Figure 4. Algorithm for preprocessing electrocardiogram (ECG) and estimating heart rate. FIR: finite impulse response; IIR: infinite impulse response.



Validation

To measure the surface resistance of the silver compound–coated electrode, an EC-80P (NAPSON KOREA) surface resistance meter was used. Surface resistance was measured for samples either treated or not treated with silver compound coating. In the case of coated samples, surface resistance was measured for each of the coated and back surfaces. When measuring sheet resistance, a load of 68 Pa was applied to the textile electrode to ensure contact was made with the probe of the sheet resistance meter. Since the result may vary for each measurement, three measuring locations were randomly selected for each sample, and then the average surface resistance was calculated by repeating measurements for each location 10 times.

To evaluate the heart rate measurement performance of the waistband-wearable device, it was simultaneously compared to a commercial system used to obtain ECGs, and then heart rate estimations were compared and evaluated. The MP150 and ECG100C systems (Biopac Systems Co Ltd) were used for obtaining reference ECGs, and these were obtained through lead I by attaching electrodes under the right and left clavicle bones. The waistband-type ECG system was worn naturally under the navel, and signals were recorded simultaneously with the MP150. Acknowledge 4.3 was used to display and store ECGs obtained from the MP150 on a computer, and the in-house smartphone app was used to store data obtained from the waistband ECG system. Signals were measured for at least 5 minutes each in the standing, sitting, and prone positions to evaluate measurement performance during various postures. The ECG signal was sampled at 1 kHz for the commercial systems and 125 Hz for the manufactured systems. This study was approved by the institutional review board (IRB) of Chonnam National University, South Korea (IRB No.1040198-200609-HR-061-02). The correspondence between

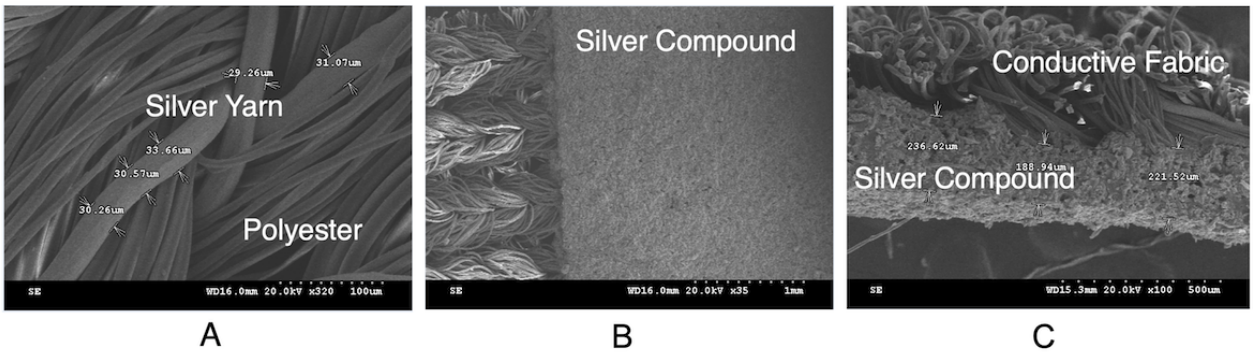
heart rates measured by the reference ECG system and the developed system were investigated using differences in the mean and standard deviation, the mean absolute error (MAE), and the mean absolute percentage error (MAPE) and by Bland-Altman analysis. Comparisons between heart rates were qualitatively evaluated.

Results

Textile Electrode

Figure 5 shows scanning electron microscope images of the fabricated textile electrode. The panels show the surface of the knitted textile electrode before silver compound coating (Figure 5, A), the surface of the knitted textile electrode after silver compound coating (Figure 5, B), and the side view cross-section (Figure 5, C). In Figure 5, A, it was shown that the knitted textile electrode was composed of silver yarn of about 30 μm in diameter with several strands of polyester. Panels B and C of Figure 5 demonstrated that the silver compound was applied with a thickness of about 190 to 240 μm. Measurement of the average surface resistance of silver-coated and uncoated electrodes revealed the surface resistance of the sample without silver coating to be 202.0 (SD 43.1) Ω/sq, whereas the surface resistance of the silver-coated and back sides were 0.22 (SD 0.03 Ω/sq) and 0.17 (SD 0.02 Ω/sq), respectively. These results demonstrated that surface resistance was dramatically reduced by the silver compound coating.

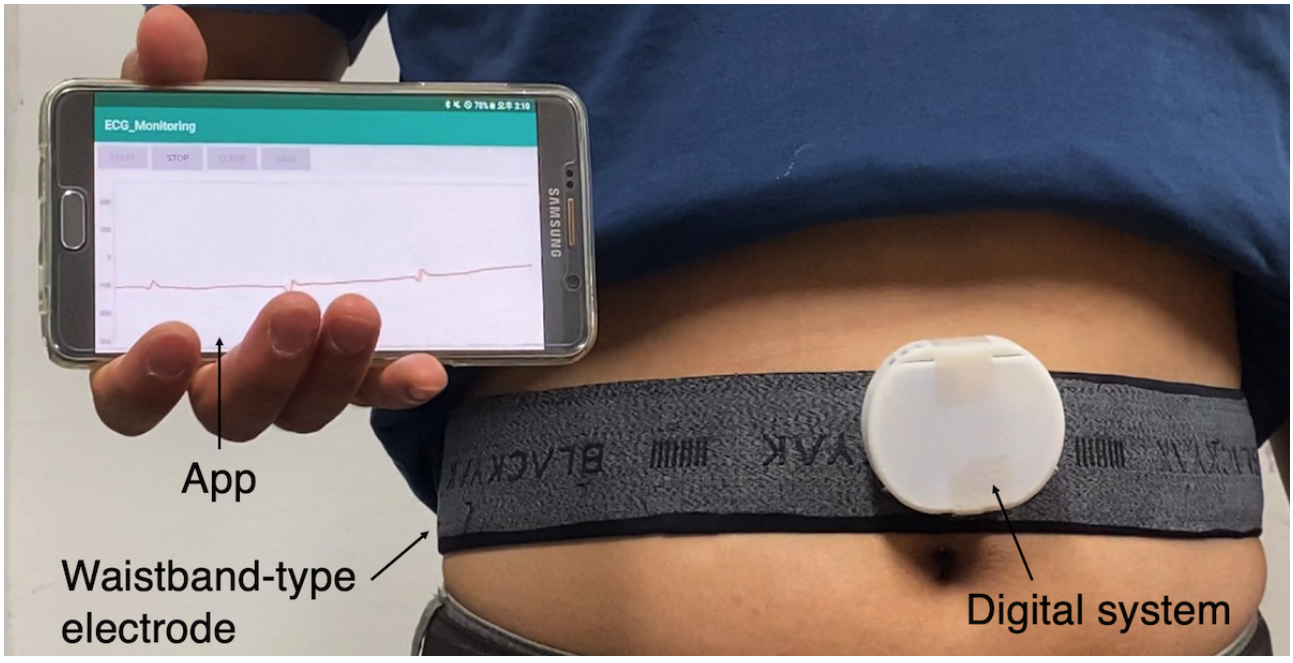
Figure 5. Scanning electron microscope images of textile electrode. A. Knitted textile electrodes without silver compound coating (320× magnification). B. Knitted textile electrodes with silver compound coating (35× magnification). C. Cross-section of knitted textile electrodes with silver compound (100× magnification).



Wireless ECG Measurement System

The size of the developed system was 53 × 45 × 12 mm (width × height × depth, including the cover) and it weighed 23 g. Figure 6 shows that the system based on our waistband-type wearable ECG electrode can measure an ECG. From the ECG waveform obtained, the QRS complex can be identified. Figure 6 displays the system components and shows that the measured ECG was successfully transmitted to a smartphone in real time. Feedback from study subjects with regard to the comfort of wearing a waistband-type wearable device elicited responses typified by “the same fit as normal pants” and “no awareness of wearing a separate sensor.”

Figure 6. Developed waistband-type wearable electrocardiogram (ECG) monitoring system, including waistband-type ECG electrode, acquisition system, and smartphone app.



Heart Rate Estimation

Figure 7 shows an example of changes to the ECG waveform based on signal processing for heart rate estimation. Figure 7, A, shows the signal received from the waistband system, in which the data interval is uneven due to deviations in the data transmission time with Bluetooth communication. Figure 7, B, shows that the ECG waveform after 1-kHz interpolation becomes uniform in interval. Figure 7, C, shows the ECG after 0.5-to-35-Hz band-pass filtering, and here the QRS waveform becomes clearer after filtering. Figure 7, D, shows the QRS complex detected by the Pan-Tompkins QRS detection algorithm (red circle). In some cases, the RRI calculated from beat loss or waveform distortion may have a value far outside the normal heartbeat interval range (eg, >1500 milliseconds). However, after correction, all the RRIs in the abnormal range disappeared and were corrected to values within the normal range. Figure 8 shows the results of heart rate estimation before and after applying the MAF. Panels A, C, and E of Figure 8 are heart rate changes in the standing, sitting, and supine positions, respectively, before using the MAF; these plots show very rapid changes, such as high-frequency noise caused by minute errors in beat intervals. On the other hand, panels B, D, and F of Figure 8 are heart rate changes after the MAF was applied in the standing, sitting, and supine positions, respectively; these plots have a trend similar to that seen when the heart rate changes gradually over time. Also seen in Figure 8, the red line is the heart rate change measured by the standard ECG measuring device MP150, and the blue line is the heart rate change measured by the developed device. From this result, it can be qualitatively confirmed that the developed device tracks the heart rate change accurately. The mean error, MAE, and MAPE of heart rates calculated from the ECG signal obtained through the developed system and the reference ECG signal are shown in Table 3. From this, it can be seen that the developed system has an MAE of 1.85 bpm in the standing position, 2.39 bpm in the sitting position, and 1.81 bpm in the lying position compared to the reference system. MAPEs were 1.80%, 2.84%, and 2.48% in the standing, sitting, and lying positions, respectively, which is acceptable and within the 10% validity criterion used in a recent study [26], as well as within the 5% validity criterion, which has been strictly applied [27]. Bland-Altman analysis (Figure 9 and Table 3) revealed that most of the mean differences between heart rates measured by both systems were within limits of agreement (LoA), as represented by dashed lines in the figure. LoA analysis revealed that for 95% of cases, the heart rate measured by the developed system will be from 0.95 to 1.04, from 0.92 to 1.07, and from 0.93 to 1.07 times the reference system in the standing, sitting, and lying positions, respectively. This indicates that heart rates measured by the developed system may differ from the reference system heart rate by 5% to 8% below to 4% to 7% above. In addition, Pearson correlation coefficients between the reference and developed systems were 0.84, 0.82, and 0.65 in the standing, sitting, and lying conditions, respectively.

Figure 7. Electrocardiogram waveform modified by signal processing procedure. A. Signal obtained via Bluetooth communication. B. Signal interpolated with 1-kHz sampling frequency. C. Signal filtered with 0.5-35-Hz finite impulse response band-pass filter. D. Detected QRS complex (red circle). ADC: analog-to-digital conversion.

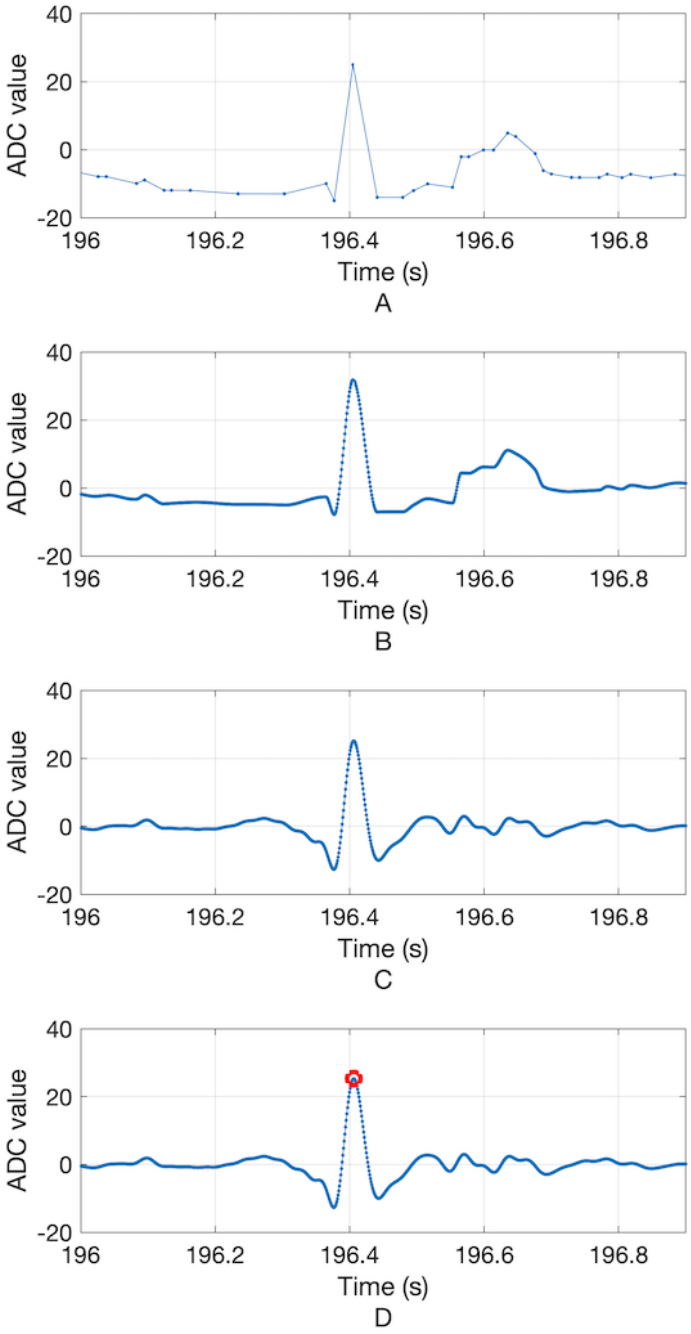


Figure 8. Bland-Altman and Pearson correlation plots. Bland-Altman plot (A) and Pearson correlation plot (B) in standing position. Bland-Altman plot (C) and Pearson correlation plot (D) in sitting position. Bland-Altman plot (E) and Pearson correlation plot (F) in lying position. bpm: beats per minute; ECG: electrocardiogram.

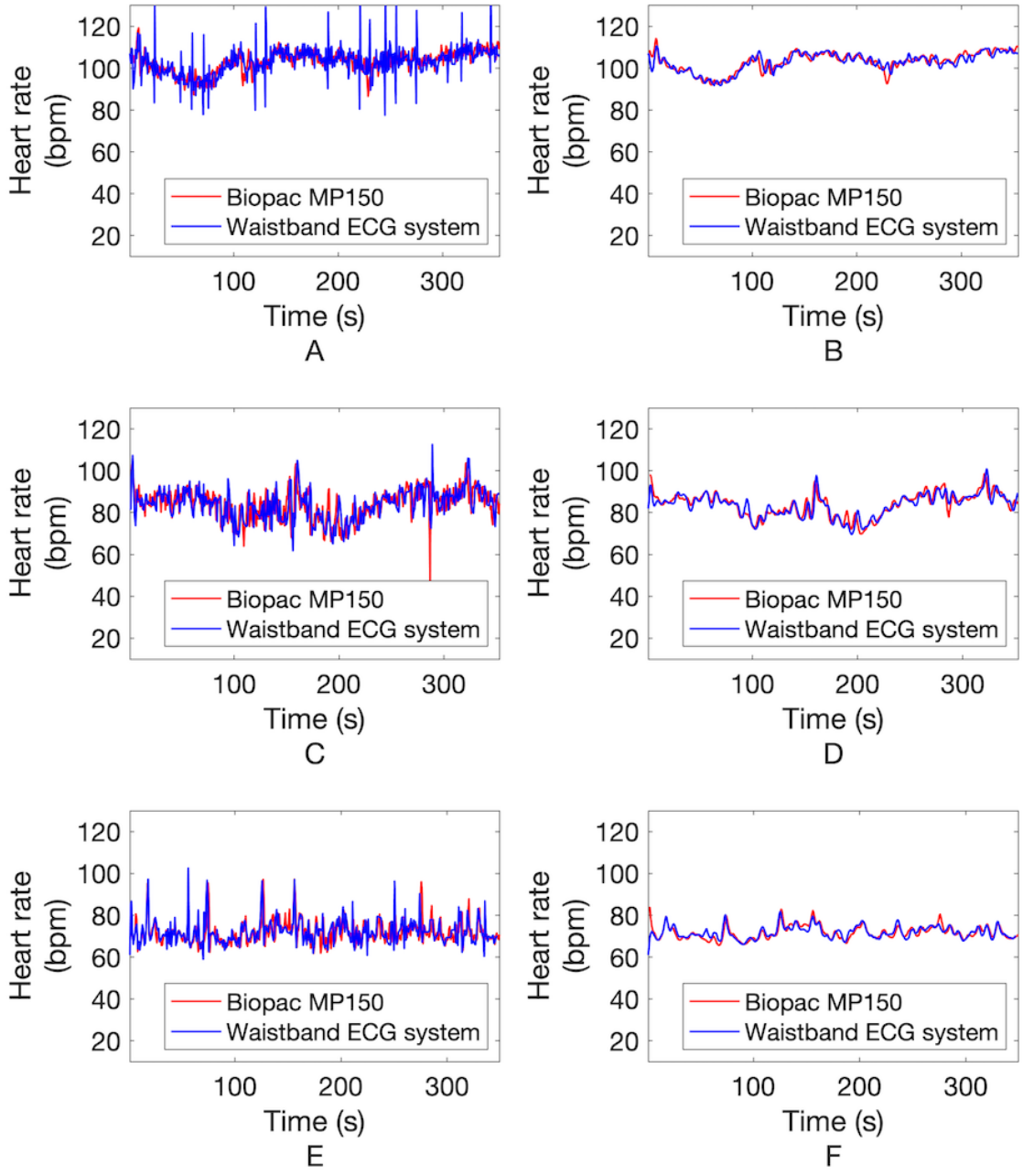


Figure 9. Heart rate change by applying the moving average filter (MAF). A. In standing position without MAF. B. In standing position with MAF. C. In sitting position without MAF. D. In sitting position with MAF. E. In supine position without MAF. F. In supine position with MAF. bpm: beats per minute; HR: heart rate; R: Pearson correlation coefficient.

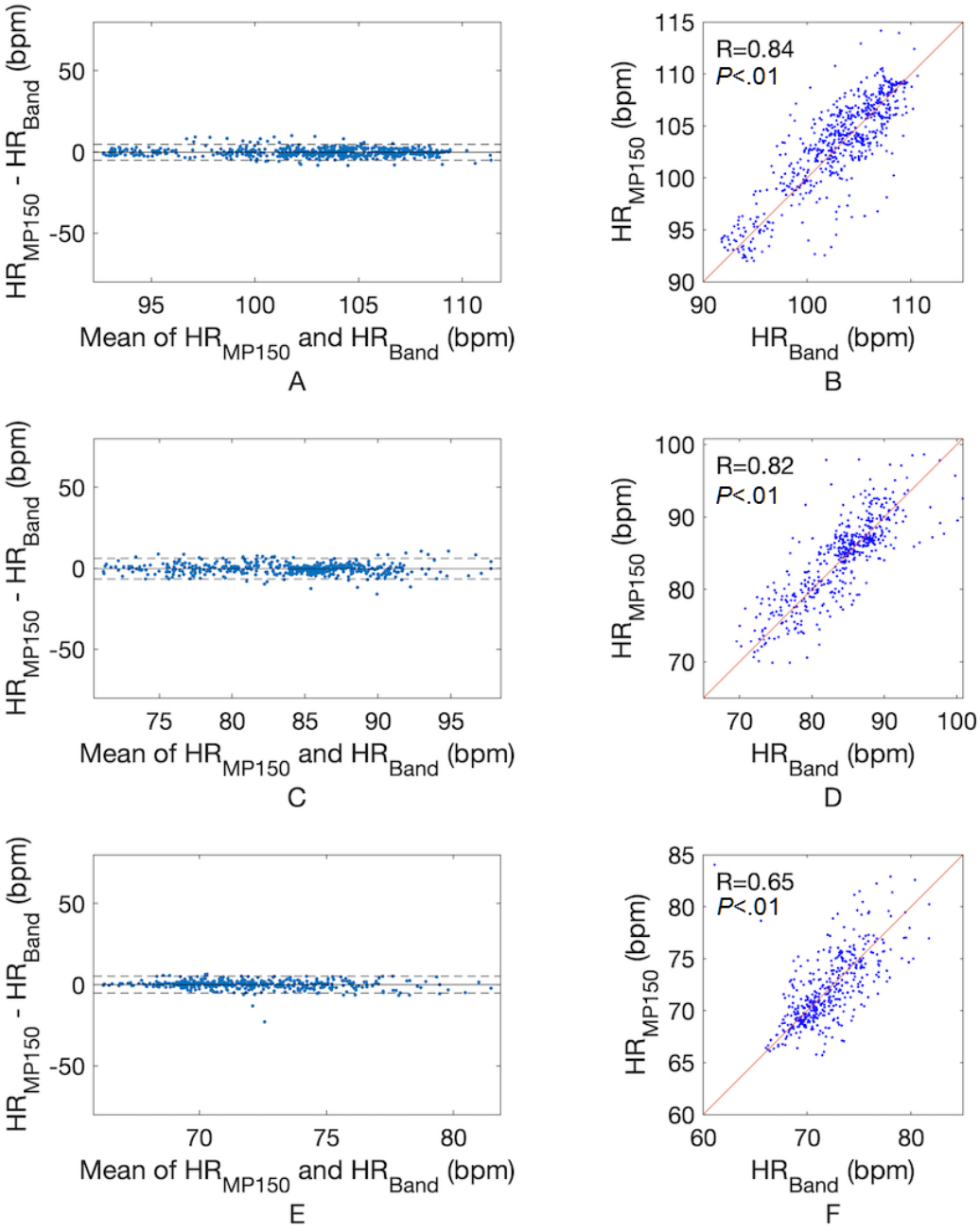


Table 3. Heart rate descriptive statistics, device error statistics, and Bland-Altman analyses comparing Biopac MP150 and waistband-type wireless electrocardiogram measurement systems by condition.

Position and device	Observations, n	Heart rate (bpm[a]), mean (SD)	Device error			Bland-Altman analysis, LoA[b]	
			MAE[c]	MAPE[d], %	Mean error (SD)	Lower	Upper
Standing							
Biopac MP150	607	103.3 (4.5)	1.85	1.80	0.23 (2.47)	–5.08	4.61
Waistband	607	103.1 (4.3)					
Sitting							
Biopac MP150	491	84.2 (5.6)	2.39	2.84	0.25 (3.27)	–6.66	6.16
Waistband	491	84.0 (5.4)					
Supine							
Biopac MP150	416	72.0 (3.3)	1.81	2.48	–0.03 (2.65)	–5.16	5.22
Waistband	416	72.0 (3.0)					

[a]bpm: beats per minute.

[b]LoA: limits of agreement.

[c]MAE: mean absolute error.

[d]MAPE: mean absolute percentage error.

Discussion

Principal Findings

In this study, a novel ECG device based on a waistband-type electrode and wireless ECG system was developed to avoid some of the disadvantages of conventional wearable devices, and its performance was verified by comparing it with a commercial device. The textile electrode was fabricated with a structure in which conductive and elastic yarns, such as polyurethane, were knitted together to provide both conductive and stretchable characteristics. The results show that this approach allowed development of an electrode that can measure an ECG and has a degree of extensibility that does not cause inconvenience during use. In addition, the impedance of the electrode was significantly reduced through silver compound coating. For ECG measurement and heart rate monitoring, representative features of the electrical signal were observed, and the MAPE of the developed system was less than 3% in heart rate measurement. This is better than the MAPEs of commercially available heart rate monitors, such as the Apple Watch or Fitbit, as reported in a previous study [26]. These results suggest that the developed system can be reliably used to obtain an ECG and to monitor heart rate. However, in real-time transmission and recording, transmission delay or loss may occur depending on the state of the communication channel. Therefore, in the future, there is a need to secure a more reliable wireless communication channel and an error correction technology for optimization. In terms of wearer comfort, the developed device has shown the possibility that such designs can be user friendly and not cause discomfort to the user. However, factors such as fit and comfort may eventually vary depending on the packaging and design of the device and, thus, need to be verified through usability evaluation after development of the final versions.

Despite the results of the study showing the potential for greatly improved user convenience, the device currently has the following limitations, which also provide directions for improvement. Firstly, the system proposed in this study was not evaluated among people with a range of physical and physiological characteristics. Therefore, for generalizing the proposed system, a follow-up study is required that would include a group of subjects with characteristics that may affect the sensitivity of the sensor, such as varied body shapes and obesity. Secondly, this study only verified the QRS detection performance with the aim of evaluating the performance of the newly proposed ECG measurement system at an early stage; however, for clinical use in the future, the measurement performance of more sophisticated ECG waveforms, such as the PR interval, QT interval, corrected QT interval, and ST segment, should be verified. Thirdly, further verification through user satisfaction surveys are also required to verify the *user-friendliness* and *comfort* characteristics that are emphasized as strengths in this study. Technically, in the case of a clothing-type ECG measurement technology using textiles, the electrode is vulnerable to motion artifacts because its contact with skin does not use an adhesive. Moreover, the electrode used in this study was a dry electrode that does not use gels, and the measured signal quality can be greatly affected by individual characteristics, such as skin condition. In addition, the durability against washing, a typical problem of textile-based electrodes, has not been considered in this study; therefore, it will be necessary to contemplate approaches such as conductive and hydrophobic coatings in the future.

Some sampling was delayed or lost during the wireless data transmission, and it has been confirmed that data delay or loss usually occurs due to the Bluetooth communication. The cause of this phenomenon is not clear, but we presume it may be due to a communication channel problem, the Bluetooth version used, or the Bluetooth strategy of the smartphone. Optimizing

the amount of transmitted data should be prioritized to reduce such transmission errors, and it is thought that performance can be improved by optimizing the sampling rate or compression transmission. Even if data loss is minimized, a transmission error occurring in the vicinity of QRS can lead to serious distortion of the waveform; therefore, research on postprocessing to compensate for this loss is also necessary. From the point of view of practical use, battery life is an important feature, but optimization of battery life and power consumption was not considered in this study. Therefore, in order to commercialize or secure practical use clinically, it will be necessary to optimize the capacity of the battery used and the power consumption of the system. In this process, a low-power MCU or a low-power design based on an interrupt in firmware can be applied. In addition, it is anticipated that battery life can be maximized through use scenarios in daily life, such as intermittent measurement.

ECGs measured with the system developed here were obtained with a modified limb lead, but this approach is known to exaggerate the amplitude of the P wave or the ST segment compared to the 12-lead standard ECG [21,28]. Therefore, it is difficult to envisage the modified limb approach replacing traditional standard limb leads, and the developed system is likely to be used for heart rate estimation and rhythmic arrhythmia detection; however, it cannot be generalized as being clinically applicable through waveform analysis. In addition, while the results of this study show that the system can be used for obtaining an ECG and calculating heart rate, it has not yet been verified on multiple subjects whose physiological characteristics may be quite diverse. Therefore, the stability and applicability of the system will need to be validated through further study on subjects exhibiting diverse individual characteristics that may vary according to criteria such as sex and age.

Conclusions

In this study, we have developed a system that can measure an ECG and that can be worn in the same way as conventional clothing in terms of appearance and ease of use, without upper body compression; we also evaluated its feasibility as a heart rate monitor. The developed system was able to measure representative ECG waveform features without user awareness or reported discomfort; in addition, the system did a good job of tracking the heart rate, as confirmed independently by a commercial ECG monitoring device. The newly developed waistband-type, wearable, wireless ECG system is anticipated to be able to secure general usability through future research aimed at improving the reliability of wireless communication technology and optimizing power consumption. Since the proposed system has not yet been verified for users with various physical and physiological characteristics, it needs to be generalized through follow-up experiments involving more diverse subjects. Moreover, it will also be necessary to confirm the convenience of the proposed system through quantitative evaluation of user satisfaction. Finally, performance verification in terms of more detailed ECG waveform characteristic feature analysis is also required for expansion to a clinical ECG measurement system.

Acknowledgments

This research was supported by two grants from the Basic Science Research Program through the National Research Foundation of Korea (NRF), one funded by the Ministry of Education (grant NRF- 2018R1D1A3B07046442) and one funded by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, Ministry of Science, ICT and Future Planning (grant NRF-2016M 3A9F1941328), Republic of Korea.

Conflicts of Interest

None declared.

References

1. Cardiovascular diseases (CVDs). World Health Organization. 2017 May 17. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2021-04-18]
2. Bouchardy J, Therrien J, Pilote L, Ionescu-Ittu R, Martucci G, Bottega N, et al. Atrial arrhythmias in adults with congenital heart disease. Circulation 2009 Oct 27;120(17):1679-1686. [doi: 10.1161/circulationaha.109.866319]
3. Srinivasan NT, Schilling RJ. Sudden cardiac death and arrhythmias. Arrhythm Electrophysiol Rev 2018 Jul 01;7(2):111-117. [doi: 10.15420/aer.2018:15:2]
4. Huikuri HV, Castellanos A, Myerburg RJ. Sudden death due to cardiac arrhythmias. N Engl J Med 2001 Nov 15;345(20):1473-1482. [doi: 10.1056/nejmra000650]
5. Chen M, Ma Y, Li Y, Wu D, Zhang Y, Youn C. Wearable 2.0: Enabling human-cloud integration in next generation healthcare systems. IEEE Commun Mag 2017 Jan;55(1):54-61. [doi: 10.1109/mcom.2017.1600410cm]
6. Cho G, Jeong K, Paik MJ, Kwun Y, Sung M. Performance evaluation of textile-based electrodes and motion sensors for smart clothing. IEEE Sens J 2011 Dec;11(12):3183-3193. [doi: 10.1109/jsen.2011.2167508]
7. Cho H, Koo H, Yang J, Lee K, Kim S, Lee J, et al. Effect of the configuration of contact type textile electrode on the performance of heart activity signal acquisition for smart healthcare. Emot Sci 2018 Dec 31;21(4):63-76. [doi: 10.14695/kjsos.2018.21.4.63]

8. Dabby N, Aleksov A, Lewallen E, Oster S, Fygenson R, Lathrop B, et al. A scalable process for manufacturing integrated, washable smart garments applied to heart rate monitoring. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: 2017 ACM International Symposium on Wearable Computers; September 11-15, 2017; Maui, HI p. 38-41. [doi: 10.1145/3123021.3123045]
9. Despang H, Netz S, Heinig A, Holland H, Fischer W. Wireless long-term ECG integrated into clothing. Biomed Tech (Berl) 2008 Dec;53(6):270-278. [doi: 10.1515/BMT.2008.043] [Medline: 19037869]
10. Hu L, Yang J, Chen M, Qian Y, Rodrigues JJ. SCAI-SVSC: Smart clothing for effective interaction with a sustainable vital sign collection. Future Gener Comput Syst 2018 Sep;86:329-338. [doi: 10.1016/j.future.2018.03.042]
11. Koo HR, Jeon DJ, Lee JH. Research on heart rate sensing clothing design for seniors based on universal fashion. Fashion Text Res J 2017 Dec 30;19(6):692-700. [doi: 10.5805/sfti.2017.19.6.692]
12. Koo HR, Lee Y, Gi S, Lee SP, Kim KN, Kang SJ, et al. Effect of module design for a garment-type heart activity monitoring wearable system based on non-contact type sensing. J Korean Soc Clothing Text 2015 Jun 30;39(3):369. [doi: 10.5850/jksct.2015.39.3.369]
13. Paiva A, Vieira D, Joana C, Carvalho H, Providência B. Design of a smart garment for cycling. In: Proceedings of the International Conference on Innovation, Engineering and Entrepreneurship (HELIX 2018). Cham, Switzerland: Springer; 2018 Presented at: International Conference on Innovation, Engineering and Entrepreneurship (HELIX 2018); June 27-29, 2018; Guimarães, Portugal p. 229-235. [doi: 10.1007/978-3-319-91334-6_32]
14. Scataglini S, Andreoni G, Gallant J. Smart clothing design issues in military applications. In: Proceedings of the International Conference on Applied Human Factors and Ergonomics. Cham, Switzerland: Springer; 2018 Presented at: International Conference on Applied Human Factors and Ergonomics; July 21-25, 2018; Orlando, FL p. 158-168. [doi: 10.1007/978-3-319-94619-1_15]
15. Shen C, Huang T, Hsu P, Ko Y, Chen F, Wang W, et al. Respiratory rate estimation by using ECG, impedance, and motion sensing in smart clothing. J Med Biol Eng 2017;37(6):826-842 [FREE Full text] [doi: 10.1007/s40846-017-0247-z] [Medline: 30220900]
16. Sundaram P, Basker N, Natrayan L. Smart Clothes with Bio-sensors for ECG Monitoring. Int J Innov Technol Exploring Eng 2019;8(4):298-301 [FREE Full text]
17. Takahashi K, Suzuki K. An ECG monitoring system through flexible clothes with elastic material. In: Proceedings of the 17th International Conference on E-health Networking, Application & Services (HealthCom). 2015 Presented at: 17th International Conference on E-health Networking, Application & Services (HealthCom); October 14-17, 2015; Boston, MA p. 305-310. [doi: 10.1109/healthcom.2015.7454516]
18. Nemati E, Deen M, Mondal T. A wireless wearable ECG sensor for long-term applications. IEEE Commun Mag 2012 Jan;50(1):36-43. [doi: 10.1109/mcom.2012.6122530]
19. Gi SO, Lee YJ, Koo HR, Khang SA, Park HJ, Kim KS, et al. An analysis on the effect of the shape features of the textile electrode on the non-contact type of sensing of cardiac activity based on the magnetic-induced conductivity principle. Trans Korean Inst Electr Eng 2013 Jun 01;62(6):803-810. [doi: 10.5370/kiee.2013.62.6.803]
20. John R, Sivaraman J. Effects of sinus rhythm on atrial ECG components using modified limb lead system. In: Proceedings of the 4th International Conference on Signal Processing, Computing and Control. New York, NY: IEEE; 2017 Presented at: 4th International Conference on Signal Processing, Computing and Control; September 21-23, 2017; Solan, India p. 527-530. [doi: 10.1109/ispcc.2017.8269735]
21. Sivaraman J, Uma G, Venkatesan S, Umapathy M, Dhandapani VE. Normal limits of ECG measurements related to atrial activity using a modified limb lead system. Anatol J Cardiol 2015 Jan;15(1):2-6 [FREE Full text] [doi: 10.5152/akd.2014.5155] [Medline: 25179878]
22. Size Korea survey. Size Korea. 2015. URL: https://sizekorea.kr/measurement-data/body [accessed 2021-01-01]
23. MPAndroidChart. GitHub. URL: https://github.com/PhilJay/MPAndroidChart [accessed 2021-08-27]
24. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;BME-32(3):230-236. [doi: 10.1109/tbme.1985.325532]
25. Morelli D, Rossi A, Cairo M, Clifton DA. Analysis of the impact of interpolation methods of missing RR-intervals caused by motion artifacts on HRV features estimations. Sensors (Basel) 2019 Jul 18;19(14):3163 [FREE Full text] [doi: 10.3390/s19143163] [Medline: 31323850]
26. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
27. ANSI/CTA Standard: Physical Activity Monitoring for Heart Rate (ANSI/CTA-2065). Arlington, VA: Consumer Technology Association; 2018 Feb. URL: https://shop.cta.tech/products/physical-activity-monitoring-for-heart-rate [accessed 2021-05-04]
28. Sheppard JP, Barker TA, Ranasinghe AM, Clutton-Brock TH, Frenneaux MP, Parkes MJ. Does modifying electrode placement of the 12 lead ECG matter in healthy subjects? Int J Cardiol 2011 Oct 20;152(2):184-191. [doi: 10.1016/j.ijcard.2010.07.013] [Medline: 20701990]

Abbreviations

AV: atrioventricular
bpm: beats per minute
ECG: electrocardiogram
IRB: institutional review board
LoA: limits of agreement
MAE: mean absolute error
MAF: moving average filter
MAPE: mean absolute percentage error
MCU: microcontroller unit
RRI: RR interval
SA: sinoatrial

Edited by L Buis, G Eysenbach; submitted 13.12.20; peer-reviewed by KC Wong; comments to author 11.03.21; revised version received 08.04.21; accepted 11.04.21; published 11.05.21.

Please cite as:
Gwon D, Cho H, Shin H
Feasibility of a Waistband-Type Wireless Wearable Electrocardiogram Monitoring System Based on a Textile Electrode: Development and Usability Study
JMIR Mhealth Uhealth 2021;9(5):e26469
URL: https://mhealth.jmir.org/2021/5/e26469
doi:10.2196/26469
PMID:33973860

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org