JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 10 (2022), Issue 9 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Reviews

Original Papers

Corrigenda and Addenda

Review

Wearable Devices in Diving: Scoping Review

Benjamin Bube[1,2], BSc, MSc; Bruno Baruque Zanón[2], PhD; Ana María Lara Palma[3], PhD; Heinrich Klocke[1], Prof Dr

[1]Faculty of Computer Science and Engineering Science, University of Applied Sciences Cologne, Gummersbach, Germany

[2]Departamento de Ingeniera Informática, Escuela Politecnica Superior, Universidad de Burgos, Burgos, Spain

[3]Departamento de Ingeniería de Organización Industrial, Escuela Politecnica Superior, Universidad de Burgos, Burgos, Spain

Corresponding Author:
Benjamin Bube, BSc, MSc
Faculty of Computer Science and Engineering Science
University of Applied Sciences Cologne
Steinmüllerallee 1
Gummersbach, 51643
Germany
Phone: 49 1601688162
Email: benjaminbube@googlemail.com

Abstract

Background: Wearables and their benefits for the safety and well-being of users have been widely studied and have had an enormous impact on the general development of these kinds of devices. Yet, the extent of research into the use and impact of wearable devices in the underwater environment is comparatively low. In the past 15 years, there has been an increased interest in research into wearables that are used underwater, as the use of such wearables has steadily grown over time. However, there has so far been no clear indication in the literature about the direction in which efforts for the design and construction of underwater wearable devices are developing. Therefore, the analysis presented in this scoping review establishes a good and powerful basis for the further development and orientation of current underwater wearables within the field.

Objective: In this scoping review, we targeted wearable devices for underwater use to make a comprehensive map of their capabilities and features and discuss the general direction of the development of underwater wearables and the orientation of research into novel prototypes of these kinds of devices.

Methods: In September 2021, we conducted an extensive search for existing literature on 4 databases and for grey literature to identify developed prototypes and early-stage products that were described and tested in water, could be worn and interacted with (eg, displays, buttons, etc), and were fully functional without external equipment. The studies were written in English, came from peer-reviewed academic sources, and were published between 2005 and 2021. We reviewed each title and abstract. The data extraction process was carried out by one author and verified by another author.

Results: In total, 36 relevant studies were included. Among these, 4 different categories were identified; 18 studies dealt primarily with safety devices, 9 dealt with underwater communication devices, 7 dealt with head-up displays, and 2 dealt with underwater human-computer interaction approaches. Although the safety devices seemed to have gained the most interest at the time of this study, a clear trend toward underwater communication wearables was identified.

Conclusions: This review sought to provide a first insight into the possibilities and challenges of the technologies that have been used in and for wearable devices that are meant for use in the underwater environment. Among these, underwater communication technologies have had the most significant influence on future developments. Moreover, a topic that has not received enough attention but should be further addressed is human-computer interaction. By developing underwater wearables that cover 2 or more of the technology categories that we identified, the extent of the benefits of such devices can be significantly increased in the future.

(JMIR Mhealth Uhealth 2022;10(9):e35727) doi:10.2196/35727

KEYWORDS

wearable device; underwater communication; head-up display; safety device; scuba diving; free diving

Introduction

Over the past few years, wearables have been widely adopted and have become tools that many people use in their daily lives [1-3]. As a result, interest in using wearables for data collection and evaluation toward a scientific purpose has also increased in many areas within the last decade, especially for the monitoring of fitness and health-related metrics [4,5].

Divers can be divided into different categories, just like mainstream wearable users. These categories include scuba divers, who tend to be recreational divers, as well as free divers, who want to stay underwater for as long as possible with 1 breath. The transition from scuba diving to technical diving is fluent. We consider technical diving to be activities that are performed beyond the depths and conditions of scuba diving. Technical divers have a clear focus on performing professional activities underwater, which mostly involves dealing with the increased demands on the equipment and with the underwater conditions.

Although divers are generally at higher risk than nondivers, far fewer studies have been conducted in this area [6]. The underwater environment is, for humans, unnatural and dangerous, which makes it particularly necessary to survey physiological factors. Such factors that relate to the pulmonary, cardiovascular, neurological, and renal systems have so far been described in detail [7-9].

Due to the significantly smaller number of people who are divers, as well as the higher demands on wearable devices in terms of water resistance and water pressure, the development of underwater wearables has been challenging. As a result of the increased pressure under water, many sensors and actuators must be treated differently than they are on land. In particular, those that provide vital sign data, such as oxygen saturation monitors or heartbeat monitors (eg, Holter monitors), must be adapted to the different underwater conditions to function smoothly. Furthermore, water represents an almost impenetrable barrier for various radio networks, such as wireless local area networks or Bluetooth networks, and the propagation of radio waves under water decreases as the frequency increases. This results in enormous hurdles, especially when networking different wearables underwater, since radio wave–based connection methods cannot be used. In addition to wired connections, acoustic and optical data transmission have primarily been investigated and recognized as useful so far [10-12].

Ongoing development and research have made it possible to propose initial prototypes, concepts, and ideas in the field of diving physiology, and wearable sensors have also been extensively investigated recently [13,14]. Therefore, using devices to collect and process diving physiology data could be helpful in minimizing underwater dangers, such as drowning, the risk of floating away, or fear. Previous studies have already been able to collect and describe in detail the individual sensors that have been used underwater [13-15]. However, none of these studies went into more detail with regard to whether and how these sensors can be combined in a portable; compact; and, if possible, networked end device. In addition, only sensors that directly relate to the health of divers have been covered in the literature so far.

To close this gap and to show a first look at the tendencies of different wearables for any kind of diver and, therefore, for general underwater use, we especially tried to answer the following questions:

- What use cases do wearables cover, besides a depth logger dive computer, and in which directions are they developing?
- Which communication technologies and seals are the most forward-looking for wearables and up to what depth have they been used and tested?
- To what extent have wearable devices been tested and what results have been attained?
- Are there important topics that have only been covered to a very small extent in scientific literature?

For the sake of completeness, individual wearables from other reviews were included, provided that they were intelligent electronic devices that could be worn on the body or on the surface of the skin and could detect, analyze, or transmit information. Dive computers that use the classic approach of a pure depth sensor were not included in this review, as we considered them to be too simple to be compared with the devices that we defined as *wearables*. Similarly, commercial dive computers were not considered in this review for two reasons. First, the details of the specifications of individual dive computers can only be compared with great effort and not in a review, since they are often tested under manufacturer-specific conditions. Second, built-in sensors and actuators are often not disclosed and would thus have required a direct comparison in a laboratory. Purely commercial studies on dive computers have been published [16-19].

Methods

The methodology used in this study was based on the approach of scoping reviews. The scoping review approach aims to present certain key concepts that have not or have only partially been reviewed so far.

In this review, we targeted wearable devices for underwater use to make a comprehensive map of their capabilities and features and discuss the general direction of the development of underwater wearables and the orientation of research and prototype designs for these kinds of devices.

For this study, a scoping review was conducted to identify and discuss the extent, scope, and nature of underwater wearable research; propose a summary of existing research; and identify gaps in the existing literature [20-22]. During the review process, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [23].

The search syntax was developed on PubMed and the Scopus database, using different word variations and combinations for the search in the “Title-Abstract” search field on PubMed and the “Article title, Abstract, Keywords” field on the Scopus

database. Each iteration of the search results was searched for 10 publications that were previously known to the authors to randomly examine the results of the search. If all 10 papers were not found, the search was repeated with a different search string. The final search formula—"wearable OR device AND (diver OR diving)"—yielded the most complete and strongest results, which contained all 10 papers. Other terms did not yield any useful results in either database and were discarded. By using the final search string, which was created to be as generic as possible, many articles that were considered for inclusion in this study were found.

The final search was conducted on September 2021 within the PubMed, Scopus, and ACM databases, and we checked Google Scholar for additional literature that may have not been covered by any of the aforementioned databases.

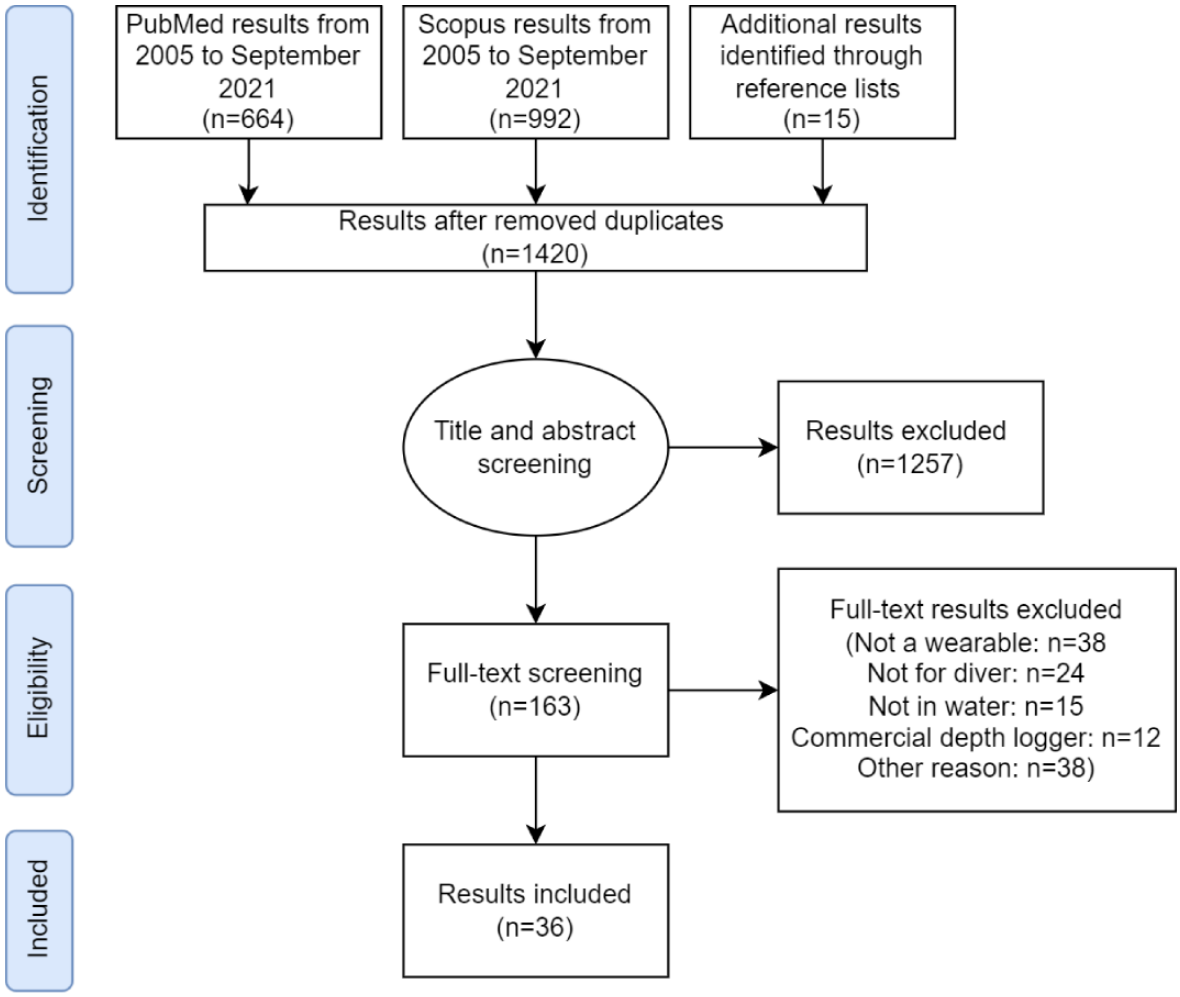
To obtain only relevant results, the search was restricted to the period after 2005. The reference lists of included articles were screened for any potentially missed papers.

A total of 2320 articles were identified by the search; PubMed returned 664 articles, Scopus returned 992 articles, and articles from additional sources were included (eg, 15 papers, which were identified from the *References* sections of the aforementioned papers and did not appear in the initial search results, were included). After excluding duplicates, a total of 1420 papers passed the initial filter and were subsequently screened based on their titles and abstracts, in terms of the objectives of this review. If a paper could not be clearly rejected or accepted based on its title or abstract because it did not match the study conditions, a full-text analysis was also carried out. To be considered as an appropriate paper, the following criteria had to be met: (1) the prototype or device was described and tested in water, (2) the device could be worn and interacted with (eg, displays, buttons, etc), (3) the device was fully functional without external equipment, and (4) the paper was written in English and published in a peer-reviewed academic source.

Wearables that could not function independently as a research object were excluded. These included individual sensors that did not function as an independent device and actuators, nonportable sensors, or systems that were not wearable devices [24-26].

After a discussion involving all authors, we decided to include 171 studies in a full-text screening. A total of 41 articles were retained for a synthesis analysis, and 5 articles from this analysis were discarded at a later stage, since they did not fully meet the inclusion criteria. The individual steps that were carried out can be seen in Figure 1.

Figure 1. Flow diagram of study selection for wearable dive computers.



A spreadsheet table was created to present the data from the individual papers. Data extraction was performed by the first author. Articles in which the individual parameters of wearable devices were not clear were examined more closely by all authors, and a joint decision was made. There was no documented instance in which a final consensus was not achieved among the four authors. The data extracted from the articles that were included in this review were the types of articles, sources, titles, study topics, study samples, housing and sealing types, depth ratings, wearables' locations on the body, implemented sensors and actuators, results, and other studies that used the same devices. In particular, the housing types and the tested depths were listed on a separate spreadsheet and displayed in a graphic. The applications of the studied wearables were compared with the respective preferred methods of wear; however, these data did not add any further value and were therefore discarded.

Results

Overview of Included Studies

Of the 36 studies retained in this review, 9 focused on underwater communication devices, 7 dealt primarily with the development of a head-up display (HUD), 2 dealt with underwater human-computer interaction possibilities, and the remaining 18 dealt with different kinds of safety devices. Theoretical measures were extracted from the wearable components' specifications, if this information was available. Herein, the *maximum depth* is the depth at which a wearable was successfully used and tested underwater, and the *construction depth* is the theoretically possible depth, which was based on the designs and commercial specifications of the components reported in the corresponding publications.

Use Cases That Wearables Cover, Besides a Depth Logger Dive Computer, and the Directions in Which They Are Developing

Safety Devices

Overview of Safety Devices

Half of the studies (18/36, 50%) dealt with divers' safety. This is nothing unusual, considering that when something dangerous happens underwater, it usually ends fatally [6]. The applied maximum test depths varied between 2.7 m and a theoretical 300 m. In this section and in the tables, both the tested and theoretical depth measures are analyzed.

Safety devices can be divided into and specified as further subcategories. The primary areas of application were vital signs (7 studies), the determination of a diver's underwater position (7 studies), breathing detection (2 studies), and cognitive functions (2 studies). Although breathing detection is a part of the *vital signs* category, this has its own section due to its relevance and importance.

Vital Signs

As shown in Table 1, of the 7 studies on vital signs, 4 collected electrocardiogram values as the subject of the study. In doing so, depths ranging from 2.7 m up to 30 m were reached. All systems could be worn regardless of their location on the body or the need for external devices. Measuring oxygen saturation, blood pressure, and heart rate is particularly relevant for free divers.

Table 1. Studies covering vital signs.

Authors	Study topic	Maximum depth (construction depth), m	Sensors and actuators	Results
Tocco et al [27]	HR[a], SV[b], and CO[c] during DA[d]	3 (90)	Miniaturized impedance cardiograph	No changes in HR, SV, and CO when compared with surface breathing and when immersed at the surface and at a 4-m depth
Tocco et al [28]	HR, SV, and CO	3 (90)	Impedance and ECG[e] recorder	Bradycardia and decrements in SV and CO
Schuster et al [29]	Measuring body temperature (core and skin) and ECG monitoring	30 (N/A[f])	ECG, temperature sensor, and Bluetooth sensor	Weak housing and problematic cables
Cibis et al [30]	Underwater monitoring of a diver's ECG signal, including an alert system that warns the diver of predefined medical emergency situations	2.7 (N/A)	ECG sensor	Showed the good accuracy of the analysis system as well as the alert system
Kuch et al [31]	Wrist-mounted apnea dive computer for the continuous plethysmography monitoring of oxygen saturation and HR	11 (200)	Transcutaneous oxygen saturation, HR, plethysmography pulse waveform, depth, time, and temperature sensors	Continuous measurement of oxygen, HR, and plethysmography pulse waves for water temperature and depth was successful
Sieber et al [32]	Measurement of blood pressure underwater	10.5 (200)	Pressure sensor and sphygmomanometer	Accurate noninvasive measurement of blood pressure underwater
Di Pumpo et al [33]	Detecting peripheral oxygen saturation for an electronic closed-circuit rebreather diver	14 (N/A)	Pulse oximeter	Detecting pulse oximetry during an immersion makes diving with a rebreather safer

[a]HR: heart rate.

[b]SV: stroke volume.

[c]CO: cardiac output.

[d]DA: dynamic apnea.

[e]ECG: electrocardiogram.

[f]N/A: not applicable.

Breathing Detection

The detection of breathing has received too little attention in the literature so far but is of elementary importance for maintaining or increasing safety underwater. To know whether a scuba diver is drowning, it is helpful to know whether the diver is still breathing. Two different approaches were successfully tested (Table 2). A precision of over 97% was achieved with one device, which "read" the intermediate pressure signals on the scuba regulator and evaluated them via a built-in algorithm. In the second study, a textile sensor was attached to a diver's chest, which expanded accordingly while the diver breathed and thus provided different values. With this system, a breathing signal can be read independently of the scuba equipment. The studies achieved depths of 25 m [34] and 30 m [35], which are acceptable for recreational diving.

Table 2. Studies covering breathing detection.

Authors	Study topic	Maximum depth (construction depth), m	Sensors and actuators	Results
Altepe et al [34]	Breathing detection device	25 (100)	2 pressure sensors	Sensitivity as high as 97.5% for 16 dives after 13.9 hours of recording
Eun et al [35]	Enhance safety and collect biometric information	30 (N/A[a])	MS5803-14BA pressure sensor (SparkFun Electronics) and respiratory sensors	Steps for entering the detailed menu should be shortened, and setting functions that were deemed to be unnecessary and dangerous (eg, rising speed warning alarm function) should be removed; diving computer usability obtained an overall average valuation of 84.7%

[a]N/A: not applicable.

Underwater Posture Determination of a Diver

The determination of a diver's position under water has been pursued by many studies via multiple approaches (Table 3). In this review, a clear tendency toward a specific solution for recording the general position of a diver was not discernible, since the technologies described covered very different approaches. However, an automated buoyancy vest is particularly suitable for scuba and technical divers, as commercial products have already shown [36,37]. Additionally, a posture determination approach that is suitable for free divers can be carried out by means of a depth sensor and an inertial measurement unit (IMU) [38]. Other studies have also pursued the possibility of recording the unconscious behaviors of a diver via a camera (attached to the diver's back) or have determined the position of a diver via a GPS or handheld sonar device.

Table 3. Studies covering underwater posture determination.

Authors	Study topic	Maximum depth (construction depth), m	Sensors and actuators	Results
Valenko et al [36]	Automatic buoyancy control	30 (80)	Pressure sensors (water and first-stage sensors), 3D accelerometer, and pneumatic valves	Initial correlations between the real dives and the simulated dive results were satisfactory.
Allotta et al [37]	Increase the safety of divers (aimed to detect the occurrence of too fast, possibly uncontrolled ascents of the diver)	300 (N/A[a])	The *SARIS* system (pressure sensors were used but not described)	The proposed application of the *SARIS* system seems feasible.
Beluso et al [39]	Automatically collect and average pressure data	4.4 (30)	Pressure sensor, display, and magnetic induction switch	Sunlight and temperature affect the pressure sensor; therefore, misleading results for the depth were obtained.
Groh et al [38]	Underwater pose determination	N/A	3-axis accelerometer, 3-axis gyroscope, and camera	The system could analyze poses and fin kicks in real time.
Hirose et al [40]	Enhance the diving experience by recording users' unconscious behaviors	N/A	Camera and wire transmission to the diver	The camera can capture the diver fully and even other diving members.
Kuch et al [41]	Accurate and affordable georeferencing for diver	N/A (300)	GPS, pressure sensor, and display	The authors reported an accuracy of <5 m.
McGrane et al [42]	Determine whether a handheld sonar device reduces the mean time for locating a missing diver	9 (N/A)	Mark Track sonar dive equipment (RJE International Inc)	The handheld sonar significantly reduces the mean duration for locating a missing diver.

[a]N/A: not applicable.

Cognitive Functions

To move in a strange and hostile environment, such as an underwater environment, intact cognitive functions are required. Although this topic is extremely relevant, only 2 systems that could function independently were identified (Table 4). In one study [43], the effects of cold water and cognitive impairment were recorded, with a significant increase of 111.7% in critical flicker fusion frequency values. The other study [44] could only determine a reduced performance at a depth of 20 m when processing the Stroop test; at a depth of 5 m or on land, no changes were found.

Table 4. Studies covering cognitive functions.

Authors	Study topic	Maximum depth, m	Sensors and actuators	Results
Piispanen et al [43]	CFFF[a] test	45	Display and flickering light-emitting diode light	Increase of 111.7% in CFFF values when compared to those in predives; skin temperature dropped by 0.48 °C
Steinberg and Doppelmayr [44]	Stroop test, Number/Letter test, 2-back test, and a simple reaction time test	20	Heart rate sensor and pressure tank air stored with Galileo Sol (Johnson Outdoors Inc)	Several findings and results

[a]CFFF: critical flicker fusion frequency.

Head-Mounted Display Devices

HUDs are difficult to manufacture for mainstream use and are difficult to design in an appealing way, as the aesthetics are primarily determined by the manufacturer based on functionality. As a result, only large manufacturers of electronic devices could claim this market for themselves; they developed HUDs for everyday use, such as Microsoft HoloLens, Google Glass, or Intel Vaunt, but even the production of these HUDs has partially stopped for the time being due to their lack of acceptance as mainstream devices [45,46], even though they

are currently offered to help people carry out work activities. However, since functionality is clearly in the foreground of diving and a diving mask or a full-face mask is used in diving anyway, the acceptance of HUDs among divers is substantially higher. As can be seen in Table 5, there are relatively many studies devoted to a HUD for divers. Due to the extensive integration of HUD technology into the masks themselves and the associated balanced pressure, most of the HUDs for diving can withstand a significantly greater depth or pressure without any problems. This is of particular benefit to technical divers, as their diving environment is the most likely to require HUD use. Of the 7 studies on HUDs, only 2 used a see-through mounted HUD. Only 1 of the 2 HUDs can be attached to a conventional mask and thus can be potentially used in free diving. Since HUDs are often integrated into full-face masks, these HUDs are not subject to the classic challenges of wearable devices and can instead be placed in masks without waterproof housing and without direct seals. As a result, these HUDs can reach significantly greater depths than those reached by HUDs mounted outside of masks.

Table 5. Studies covering head-mounted display devices.

Authors	Type of HUD[a]	Mounting	Dive mode	Maximum depth (construction depth), m	Sensors and actuators
Koss and Sieber [47]	Not see-through	Mounted outside of a mask	Rebreather diving	300 (N/A[b])	HUD, 3 pO_2[c] sensors, depth sensor, time sensor, and decompression obligation sensor
Sieber et al [48]	Not see-through	Mounted inside of a full-face mask	Scuba diving, surface-supplied gas diving, and rebreather diving	45 (100)	Full-color display, depth sensor, tilt-compensated compass, and tank pressure sensor
Gallagher et al [49]	Not see-through	Mounted outside of a mask	Military combat diving	N/A	Microdisplay, optical lens, electronic compass, depth sensor, microprocessor, associated electronics, and battery
Gallagher and Manley [50]	See-through	Mounted inside of a mask	Scuba diving, surface-supplied gas diving, and rebreather diving	9 (N/A)	Depth sensor, compass, light-emitting diodes, and HUD
Manley et al [51]	See-through	Diving helmet	Military combat diving	12 (N/A)	HUD
Koss and Sieber [52]	Not see-through	Mounted outside of a mask	All (copies the dive computer screen)	95 (300)	Bluetooth sensor, pressure sensor, display, buttons, and pO_2 sensor
Sieber et al [53]	Not see-through	Mounted outside of a mask	Rebreather diving	130 (N/A)	Infrared receiver, 3-axis IMU[d], pressure sensor, tank pressure sensors, galvanic pO_2 sensors, display, and buttons

[a]HUD: head-up display.

[b]N/A: not applicable.

[c]pO_2: partial pressure of oxygen.

[d]IMU: inertial measurement unit.

Underwater Communication Devices

Underwater communication is the most important aspect of a wearable for the Internet of Underwater Things (IoUT) [12]. Without a wireless connection between each device and to the internet, the IoUT would not be able to establish itself. This is why wireless connectivity is of particular importance to the future development and establishment of the IoUT.

Due to the complexities and sizes of the modems for wireless underwater transmission, there is so far only a handful that has been successfully implemented and tested in a wearable. As seen in Table 6, two such wearables fall back on a 2-part solution in which the transmitter or receiver is attached to the back, and the diver simply connects a wearable to the device on their back [54,55]. Furthermore, apart from those in a study by Bube et al [56], none of the wearables can reach a range of more than 20 m, which is not sufficient for meaningful use. Additionally, the data rate also decreases considerably as the communication range increases.

Table 6. Studies covering underwater communication devices.

Authors	Communication technology	Data rate	Maximum range, m	Power consumption, W	Depth, m	Sensors and actuators
Hussein et al [57]	Light	N/A[a]	A few meters	N/A	1	N/A
Kohlsdorf et al [58]	Acoustic	N/A	18 (direct positioning)	>5 and <10	N/A	Hydrophones, speaker, keyboard, and display
Cardia et al [54]	Acoustic	N/A	N/A	N/A	N/A	N/A
Anjangi et al [55]	Acoustic	N/A	>50	N/A	6.6	Beacon (GPS), acoustic communication, MS5837-30BA sensor (TE Connectivity), and pressure transducer
Chen et al [59]	Optical	500 kB/s	20	<10	30	Camera, photoelectric sensor, audio acquisition, and display
Katzschmann et al [60]	Acoustic	20 bytes/s	15	N/A	18	Acoustic transducer, depth sensor, and IMU[b]
Kuch et al [61]	GPS/GSM[c] cable	N/A	N/A	N/A	16	Pressure sensor, tank pressure sensor, GPS/GSM, and display
Bube et al [56]	Acoustic	64 bytes/s	200	2.6	250	Pressure sensor, RTC[d], acoustic modem, temperature sensor, heartbeat sensor, and display
Navea and Claveria [62]	Light	4 kB/s	7	N/A	1.5	Light sensors, earphones, and phototransistors

[a]N/A: not applicable.

[b]IMU: inertial measurement unit.

[c]GSM: Global System for Mobile Communications.

[d]RTC: real-time clock.

Human-Computer Interaction Approaches

The interaction with a conventional dive computer usually takes place via various buttons that are sealed against the ambient pressure under water. Although there is great potential for improvement in this area, only 2 of the papers dealt with the topic of interaction, as seen in Table 7. For this purpose, both the implementation of a touch screen that was insensitive to water pressure and the implementation of interaction via tilting the device for input were tested. For both variants, the advantage over button-based interaction stands out. Furthermore, there is no need for a physical connection to the outside of the housing, which always represents a potential weak point.

In both studies, good results were achieved under all conditions, which makes conducting further tests in this direction appear sensible. A comparison between the two interaction options and those for interacting with a classic dive computer via buttons makes sense.

Table 7. Studies covering human-computer interaction approaches.

Authors	Interaction type	Mounting type	Maximum depth, m	Sensors and actuators
Lee and Jun [63]	Touch screen	Wrist	50	Temperature, water pressure, and direction sensors
Čejka et al [64]	Tilting for underwater typing	Handheld	5	Samsung S8 sensors

The Depths at Which the Wearables Were Used and the Most Forward-looking Seals

Particular attention should be paid to the implementation of housing in a large number of different studies, as this is currently one of the greatest hurdles for the development of new and innovative ideas for the IoUT. As can be seen in Table 8, almost every housing type was used for different study designs. Nevertheless, a clear tendency in the choices of the primarily used housing types was seen. The most commonly reported housing type was a polymer or polymethylmethacrylate (transparent thermoplastic plastic) housing (studies: 15/36, 42%). This was followed by aluminum cases (studies: 7/36, 19%) and commercial smartphone, tablet, or bag cases (studies: 9/36, 25%). There were also instances of devices being sealed either in a diving helmet (studies: 3/36, 8%) or with a potting compound (studies: 3/36, 8%). Only 1 of the 36 (2%) studies used tempered glass for the housing. In 2 of the 36 (5%) studies, no information on the housing was given.

Table 8. Housing and sealing comparison.

Housing and sealing type and study topic	Tested depth (construction depth), m
Aluminum case	
Communication	
Kohlsdorf et al [58]	N/A[a]
Kuch et al [61]	16 (N/A)
Head-up display	
Sieber et al [48][b]	45 (N/A)
Safety device	
Valenko et al [36][b]	30 (80)
Altepe et al [34]	25 (100)
Kuch et al [41]	N/A (300)
Di Pumpo et al [33][b]	14 (N/A)
Commercial smartphone housing, tablet housing, or bag	
Communication	
Hussein et al [57]	1 (N/A)
Cardia et al [54]	N/A
Anjangi et al [55]	6.6 (N/A)
Interaction	
Čejka et al [64]	5 (N/A)
Safety device	
Beluso et al [39]	4.4 (30)
Steinberg and Doppelmayr [44]	20 (N/A)
Groh et al [38]	N/A
Schuster et al [29]	30 (N/A)
Cibis et al [30]	2.7 (N/A)
Diving helmet or mask	
Communication	
Chen et al [59]	30 (N/A)
Head-up display	
Manley et al [51]	12 (N/A)
Safety device	
Di Pumpo et al [33][b]	14 (N/A)
Polymer or polymethylmethacrylate (Lexan, acryl, Plexiglas, etc)	
Communication	
Navea and Claveria [62]	1.5 (N/A)
Katzschmann et al [60]	18 (N/A)
Bube et al [56]	250 (N/A)
Head-up display	
Koss and Sieber [47]	300 (N/A)
Sieber et al [48][b]	45 (100)
Gallagher et al [49]	N/A
Koss and Sieber [52][b]	95 (300)

Housing and sealing type and study topic	Tested depth (construction depth), m
Safety device	
Valenko et al [36][b]	30 (80)
Kuch et al [31]	11 (200)
Tocco et al [27]	3 (90)
Sieber et al [32]	10.5 (200)
Piispanen et al [43]	45 (N/A)
Tocco et al [28]	3 (90)
Allotta et al [37]	300 (N/A)
Hirose et al [40]	N/A
Potting compound	
Head-up display	
Gallagher and Manley [50][c]	9 (N/A)
Koss and Sieber [52][b]	95 (300)
Sieber et al [53]	130 (N/A)
Tempered glass	
Interaction	
Lee and Jun [63]	50 (N/A)
Not specified	
Safety device	
Eun et al [35]	30 (N/A)
McGrane et al [42]	9 (N/A)

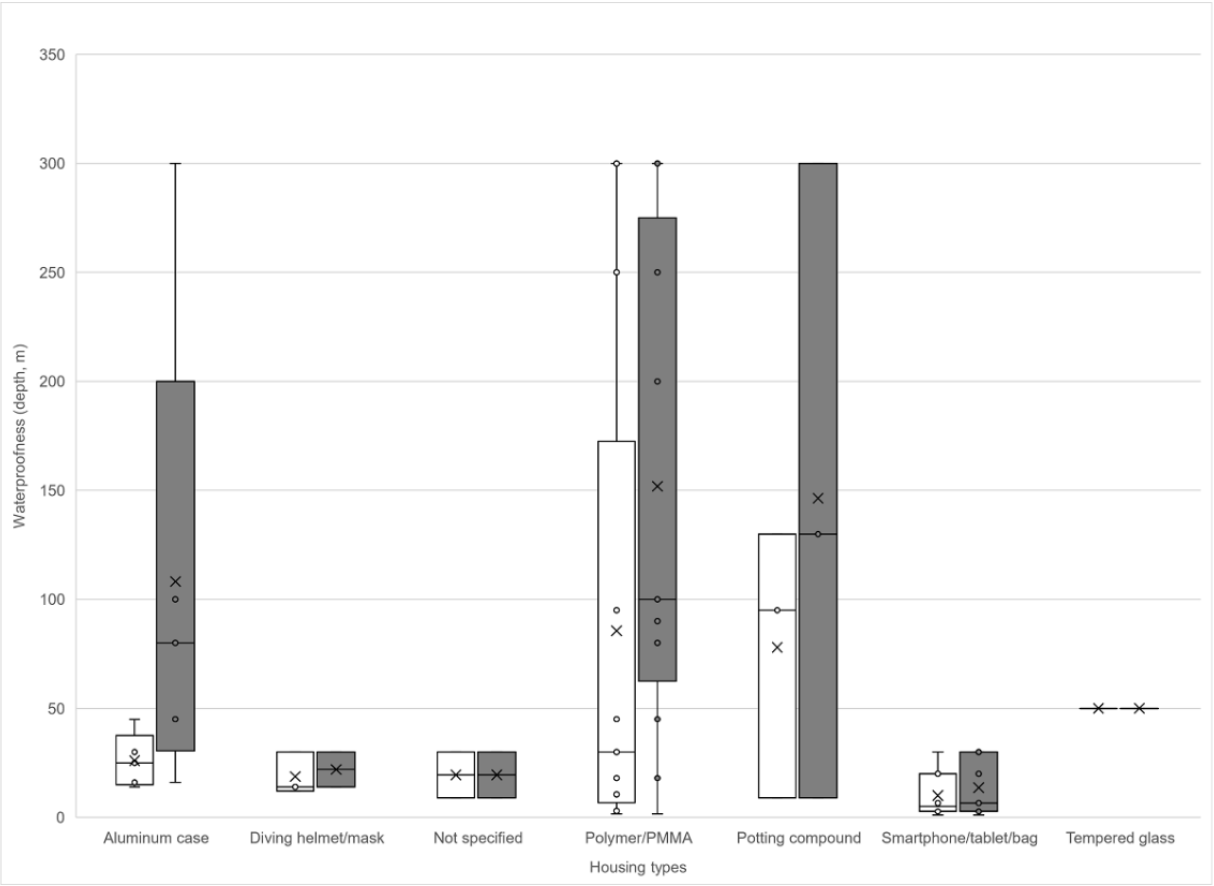
[a]N/A: not applicable.

[b]The device consists of 2 parts and is therefore listed in 2 housing and sealing categories.

[c]The depth is only 9 m because the potting compound was not applied to all components. Instead, the components had their own compartments, which is a problem with regard to sealing. The buttons were sealed with O-rings.

As expected, across all examined studies, the depth tested was well below the theoretical construction depth (Figure 2). Only the use of tempered glass was tested at the maximum specified depth. A direct comparison between housings that were made of a polymer or polymethylmethacrylate and housings that used a potting compound showed that the depths achieved by both housing types were approximately equivalent. By weighing the costs and benefits of a specific study that is to be carried out, a decision can be made between the two housing types. The most common primary cause cited against the use of a potting compound for housing was the difficulty in accessing the device (ie, for charging, programming, and interacting with the device) after pouring the compound [39]. If these challenges are overcome, the use of cast housings, including those used for underwater sensors, could prevail in the long term.

Figure 2. Housing type and waterproofness (ie, depth in meters). White bars indicate the tested and confirmed depths. Grey bars indicate the calculated or specified depths. PMMA: polymethylmethacrylate.



Discussion

Study Overview

Some of the insights that we gained from the reviewed studies were that many of the wearables examined were in the prototype stage or were only designed for a specific group of users. Safety-relevant devices received the greatest attention, and much of the technological progress in the development of underwater wearables can be attributed to their contributions to the field.

The most promising areas of development include underwater communication and human-computer interaction, as improvements in these areas will enable the entry of underwater wearables into a consumer market, which in turn can result in increased attention for such wearables in society and thus increased attention for science.

Principal Results

Safety Devices

The collection and storage of vital values via underwater wearables received the most attention, since little is known about the medical background of diving, especially among divers and free divers. However, due to the advancing developments in this area within recent years, the use of such wearable devices has made it possible to achieve initial results. By linking previously developed and functioning safety devices with underwater communication devices, the almost real-time monitoring of a diver can take place within various disciplines in the future. Possible scenarios for the use of underwater wearables for vital sign monitoring include technical diving and free diving competitions, among others, as such wearables can be used to increase the safety conditions of these activities. By specifically measuring the vital parameters of free divers in all diving competitions, a significantly larger and more meaningful database can be accessed in the future. For subsequent developments in underwater safety wearables, predictive algorithms could be developed based on a vital sign database. This algorithm could be used to warn divers about critical conditions before they occur. This approach, as a concept, has already been presented but has not yet been tested in real life [65]. To achieve this however, the necessary prototypes must be significantly further developed, so that they can be used meaningfully outside of a scientific study, preferably as a finished consumer product.

As a result of the fact that divers can move freely in all 3 dimensions, contrary to land-based deployment, location and position determination received substantially higher levels of attention. In addition, since GPSs do not work when used in water, various localization options were investigated, but they showed weaknesses in various settings, such as in caves, in water with strong currents, and under great depths [41]. Location and position determination via an integrated IMU has shown promising results and should be further investigated. The almost real-time transmission of the underwater position of a diver also has many applications, such as the monitoring of individual students at a diving school; the early detection of dangers, such

as currents and drifts; and, for free divers, the analysis of movement sequences by a trainer.

The use of breathing detection devices in combination with underwater communication devices also provides new possibilities. In addition, by using a developed respiratory sensor system, a further focus on the collection and evaluation of diaphragm contraction data from free divers can be achieved [35].

Underwater Communication

Underwater communication via multiple underwater devices and communication with people on the surface are fundamental pillars for an extensive network of wearables. However, the spread of this network is currently limited by 3 essential factors. These factors are the sizes, costs, and bandwidths of modules, which currently do not allow for any economic dissemination. Due to the constantly advancing developments in the IoUT and the underlying sensor networks, these modules will be useful in the future due to the scalability of wearables for underwater use. The uses of underwater communication wearables are diverse and can be expanded enormously.

By implementing the ahoi acoustic modem in a wearable, a significantly more compact and cheaper wearable for use underwater can be created in the future [66]. If the development of underwater wearables proceeds in the same manner as the development of mainstream wearables, underwater wearable development can focus on miniaturization and arrangement, which would benefit the current bulky modems that are used underwater [67]. By networking divers' devices with each other and with the internet, the potential of these devices was often examined and shown in regard to the IoUT [7,9].

Through the further connection of wearables to other sensors on reefs, boats, or other underwater sites (eg, shipwrecks), safety-relevant information can also be transmitted to divers regardless of vital parameters, underwater locations, navigation limitations, or currents, and appropriate warnings (eg, the sudden appearance of dangers, changes in current direction, etc) can be given to avoid accidents. An overview of the variety of safety options was provided by Jahanbakht et al [12].

As soon as underwater communication wearables can be made to be cheaper and more compact, other subaspects, such as the collection of vital parameter data, will automatically improve. The first promising step toward a more cost-effective and compact device with an acceptable communication range has already been presented [56]. A wearable device with data transmission capabilities can be manufactured commercially through consistent and further developments that are based on previous approaches. However, free divers' willingness to spend more money on unique products has yet to be investigated.

Of note, since GPSs do not work underwater, a different approach is required with regard to locating divers as well. If entry and exit points are recorded by a GPS, the underwater location of a diver can be determined via an IMU.

Algorithms that can recognize whether a diver is in danger based on movement data can also be used to make enormous progress toward locating divers and significantly increasing their safety. This approach has already been described in great detail for land-based use cases [68,69]. Vinetti et al [14] concluded that the monitoring and transmission of oxygen levels, as well as related feedback, and the most effective economical swimming techniques will have the greatest impact in the future. Ours is the first review to consider both the data monitoring aspects and data transmission aspects of underwater wearables. If these aspects are optimized, we believe that further developments for the IoUT will be made in the future and will have the greatest impact on the underwater world.

Human-Computer Interaction

Because the focus of the development of dive computers has so far been almost exclusively on computer science studies, it is not surprising that the human-computer interaction aspect has only been researched very rudimentarily so far. As a result, the number of studies that have been carried out on this subject has been very limited. However, the two identified interaction approaches showed a clear trend and the associated need for further investigations. A trend toward a design without weak points and connections to external components can be seen in the literature. In one study [63], interaction via the classic touch screen was chosen as the interaction method, and the other study [64] opted for interaction via tilting the device. As soon as the commercialization of underwater data communication becomes better established, as with mainstream wearables, human-computer interaction with regard to underwater wearables could gain importance and attention in the next years [70]. However, it is possible that insufficient attention is paid to this subject, which has been the case for the mainstream wearable market in recent years [67].

In the context of underwater wearables, whether interactions that do not require external components turn out to be better or more useful than interactions that do require such components (ie, buttons) should be further pursued and investigated. It may well be that the general paradigms of usability and user experience that are applied to land-based devices cannot be applied to underwater devices to the same extent. So far however, no studies have been carried out in this direction to our knowledge. Nevertheless, as soon as underwater data communication is offered as a commercial function of dive computers, the human-computer interaction aspect of underwater wearables could receive increased attention, since improvements in this aspect would result in completely new methods of interaction and extended functionalities that could address the interaction needs of users.

Housing and Sealing

Several lines of evidence suggest that the sealings used to protect against underwater environmental influences will continue to distinguish the development of wearables for underwater use from the development of wearables for land-based use in the future. As far as our review shows, no study has dealt with this subject before. Per the data we gathered on housing and sealing types and their respective achieved or projected depths, we assume that polymer or polymethylmethacrylate housings and cast housings will continue to be the primarily used housing types in the future [47]. Cast housings can be used to eliminate various problems, such as heat generation, programmability

issues, chargeability issues, and user-friendliness issues. Furthermore, the next step should be a careful study of the relationships among cost, aesthetics, design, and achievable depth, which have direct implications for various seals. These are particularly important with regard to whether a completely casted housing is accepted for a commercial product. Even though the individual reasons were not mentioned, a completely encapsulated housing was only used for 3 HUDs [50,52,53]. Therefore, no clear trend can be identified in this area. The reasons for not using such housings could have been poor aesthetics, poor maintainability, and dissatisfaction with the elimination of the heat generated by a device. The introduction of different colored epoxy adhesives, along with light-emitting diodes and a special form of dive computer, could certainly appeal to the market.

Comparison With Prior Work

To date, many reviews on underwater wearables have focused exclusively on the collection and evaluation of physiological and psychological parameters during diving as the primary research objective. These studies generally focused on available sensors that can be adapted for underwater use [13], dealt exclusively with the sensors and not with the entire wearable [14], or only dealt with devices that measured physiological parameters [15].

Apart from the fact that some of the papers we reviewed were published a few years ago, they mainly dealt with safety-related aspects in the field of diving. This trend was repeatedly confirmed in this review, since half of the reviewed studies (18/36, 50%) examined wearable devices that were related to safety. Furthermore, the safety-related wearable devices identified in this review were largely used in conjunction with devices from other studies. However, these other studies were excluded because the devices they used could not be used as a wearable, per the criteria outlined in the *Methods* section, or because they were published after the review period.

Ours is the first review on underwater devices, and we provide a first look at their potential and the challenges associated with their development. So far, the development areas of underwater communication and human-computer interaction for divers have not received any real attention. Furthermore, this is the first review to summarize the available diving devices that can be considered scientifically tested wearable prototypes. Commercial dive computers themselves have already been studied in terms of various parameters, such as precision in measuring depth and ergonomic performance [18,19]. A review of modern dive computers and a comparison of 47 dive computer models, which involved a comparison of their specifications, were carried out before [16,17]. By gathering data on the housing types of wearable devices along with the maximum tested depths and the theoretically calculated depths at which the devices remained functional, we were also able to show the tendencies in this area, which have not been shown before, as far as we know.

Conclusions

This scoping review shows a first comprehensive insight into the various subaspects of developed prototypes of wearable devices for underwater use. The possibilities and challenges of the reviewed technologies were considered and evaluated separately. In addition to the well-covered field of safety devices that relate to the collection of vital sign data from divers, other areas such as underwater communication between divers, as well as topics such as human-computer interaction and specialized wearables for divers, were covered for the first time. Recent research has shown that underwater communication has the most significant influence on future developments. In contrast, human-computer interaction has so far received far too little consideration. This is particularly surprising because the conditions under water are different from those on the surface. A scientific summary and overview of the housings and seals used among devices for scientific purposes should be considered in the future and on a larger scale.

In their current state, none of the devices reviewed in this study can prompt the further development of underwater wearables. The greatest future impacts will result from a combination of all of the aspects mentioned herein, with a special focus on safety and communication. The trends seen with mainstream wearables can thus be seen with underwater wearables as well, which focus primarily on sensor design, communication protocols, and data processing and analysis [67]. If these trends continue, underwater safety devices could be used to communicate with other divers and stations in the IoUT and, if necessary, immediately carry out an action. This could, for example, significantly shorten and optimize a rescue chain in an emergency. By focusing research on wearable devices for underwater use and further developing them into consumer products, such underwater networking could also be used for subareas other than safety measures or the collection of human physiology data [13]. The possible application scenarios could include the maintenance and repair of underwater structures, such as bridges or drilling platforms; the collection and evaluation of data from animals by using sensor materials; or the broad-based collection of data on submarine environments by using wearable devices underwater.

Conflicts of Interest

None declared.

References

1. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer perceptions of wearable technology devices: Retrospective review and analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]

2. Pobiruchin M, Suleder J, Zowalla R, Wiesner M. Accuracy and adoption of wearable technology used by active citizens: A marathon event field study. JMIR Mhealth Uhealth 2017 Feb 28;5(2):e24 [FREE Full text] [doi: 10.2196/mhealth.6395] [Medline: 28246070]
3. Chang CC. Exploring the usage intentions of wearable medical devices: A demonstration study. Interact J Med Res 2020 Sep 18;9(3):e19776 [FREE Full text] [doi: 10.2196/19776] [Medline: 32945778]
4. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
5. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: Comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
6. Divers Alert Network. In: Buzzacott P, Denoble PF, editors. DAN Annual Diving Report 2018 Edition: A Report on 2016 Diving Fatalities, Injuries, and Incidents. Durham, NC: Divers Alert Network; 2018.
7. Dujic Z, Breskovic T. Impact of breath holding on cardiovascular respiratory and cerebrovascular health. Sports Med 2012 Jun 01;42(6):459-472. [doi: 10.2165/11599260-000000000-00000] [Medline: 22574634]
8. Chonody JM, Webb SN, Ranzijn R, Bryan J. Working with older adults: Predictors of attitudes towards ageing in psychology and social work students, faculty, and practitioners. Aust Psychol 2020 Nov 12;49(6):374-383. [doi: 10.1111/ap.12056]
9. Gren M, Shahim P, Lautner R, Wilson DH, Andreasson U, Norgren N, et al. Blood biomarkers indicate mild neuroaxonal injury and increased amyloid β production after transient hypoxia during breath-hold diving. Brain Inj 2016;30(10):1226-1230. [doi: 10.1080/02699052.2016.1179792] [Medline: 27389622]
10. Yin H, Li Y, Xing F, Wu B, Zhou Z, Zhang W. Hybrid acoustic, wireless optical and fiber-optic underwater cellular mobile communication networks. 2019 Jan Presented at: 2018 IEEE 18th International Conference on Communication Technology (ICCT); October 8-11, 2018; Chongqing, China p. 721-726. [doi: 10.1109/icct.2018.8599957]
11. Chowdhury MZ, Hasan MK, Shahjalal M, Hossan MT, Jang YM. Optical wireless hybrid networks: Trends, opportunities, challenges, and research directions. IEEE Communications Surveys & Tutorials 2020 Jan 15;22(2):930-966 [FREE Full text] [doi: 10.1109/comst.2020.2966855]
12. Jahanbakht M, Xiang W, Hanzo L, Azghadi MR. Internet of Underwater Things and big marine data analytics—A comprehensive survey. IEEE Communications Surveys and Tutorials 2021 Jan 20;23(2):904-956. [doi: 10.1109/comst.2021.3053118]
13. Cibis T, McEwan A, Sieber A, Eskofier B, Lippmann J, Friedl K, et al. Diving into research of biomedical engineering in scuba diving. IEEE Rev Biomed Eng 2017;10:323-333. [doi: 10.1109/RBME.2017.2713300] [Medline: 28600260]
14. Vinetti G, Lopomo NF, Taboni A, Fagoni N, Ferretti G. The current use of wearable sensors to enhance safety and performance in breath-hold diving: A systematic review. Diving Hyperb Med 2020 Mar 31;50(1):54-65 [FREE Full text] [doi: 10.28920/dhm50.1.54-65] [Medline: 32187619]
15. Sieber A, L'Abbate A, Kuch B, Wagner M, Benassi A, Passera M, et al. Advanced instrumentation for research in diving and hyperbaric medicine. Undersea Hyperb Med 2010;37(5):259-269. [Medline: 20929183]
16. Azzopardi E, Sayer MDJ. A review of the technical specifications of 47 models of diving decompression computer. Underwater Technology 2010 Jul 01;29(2):63-72 [FREE Full text] [doi: 10.3723/ut.29.063]
17. Wienke BR, O'Leary TR. Modern dive computers. In: Understanding Modern Dive Computers and Operation: Protocols, Models, Tests, Data, Risk and Applications. The Capital Region of Denmark: Springer Cham; 2018:9-16.
18. Ozyigit T, Egi SM. Evaluating the ergonomic performance of dive computers. 2012 Jun Presented at: 2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP); May 16-18, 2012; Bangkok, Thailand. [doi: 10.1109/dictap.2012.6215418]
19. Elliott KH, Gaston AJ. Accuracy of depth recorders. Waterbirds 2009 Mar;32(1):183-191. [doi: 10.1675/063.032.0123]
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2007 Feb 23;8(1):19-32. [doi: 10.1080/1364557032000119616]
21. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
22. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: A scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Düking P, Achtzehn S, Holmberg HC, Sperlich B. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. Sensors (Basel) 2018 May 19;18(5):1632 [FREE Full text] [doi: 10.3390/s18051632] [Medline: 29783763]
25. Düking P, Hotho A, Holmberg HC, Fuss FK, Sperlich B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. Front Physiol 2016 Mar 09;7:71 [FREE Full text] [doi: 10.3389/fphys.2016.00071] [Medline: 27014077]

26. O'Donoghue J, Herbert J. Data management within mHealth environments: Patient sensors, mobile devices, and databases. ACM J Data Inf Qual 2012 Oct 01;4(1):1-20. [doi: 10.1145/2378016.2378021]
27. Tocco F, Crisafulli A, Marongiu E, Milia R, Kalb A, Concu A. A portable device to assess underwater changes of cardio dynamic variables by impedance cardiography. 2012 Dec 20 Presented at: First Latin-American Conference on Bioimpedance (CLABIO 2012); November 6-19, 2012; Joinville, Brazil. [doi: 10.1088/1742-6596/407/1/012026]
28. Tocco F, Marongiu E, Pinna M, Roberto S, Pusceddu M, Angius L, et al. Assessment of circulatory adjustments during underwater apnoea in elite divers by means of a portable device. Acta Physiol (Oxf) 2013 Feb;207(2):290-298. [doi: 10.1111/apha.12000] [Medline: 22978452]
29. Schuster A, Castagna O, Schmid B, Cibis T, Sieber A, Seabear GmbH. Underwater monitoring system for body temperature and ECG recordings. Underwater Technology 2017 Jul 01;34(3):135-139 [FREE Full text] [doi: 10.3723/ut.34.135]
30. Cibis T, Groh BH, Gatermann H, Leutheuser H, Eskofier BM. Wearable real-time ecg monitoring with emergency alert system for scuba diving. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:6074-6077. [doi: 10.1109/EMBC.2015.7319777] [Medline: 26737677]
31. Kuch B, Koss B, Dujic Z, Buttazzo G, Sieber A. A novel wearable apnea dive computer for continuous plethysmographic monitoring of oxygen saturation and heart rate. Diving Hyperb Med 2010 Mar;40(1):34-40. [Medline: 23111837]
32. Sieber A, Kuch B, L'abbate A, Wagner M, Dario P, Bedini R. An underwater blood pressure measuring device. Diving Hyperb Med 2008 Sep;38(3):128-134. [Medline: 22692713]
33. Di Pumpo F, Ruffino G, Malacarne P. Pulse oximeter to detect peripheral oxygen saturation in underwater rebreather ECCR diver: a preliminary study. Undersea Hyperb Med 2019;46(1):1-6. [Medline: 31154680]
34. Altepe C, Egi SM, Ozyigit T, Sinoplu DR, Marroni A, Pierleoni P. Design and validation of a breathing detection system for scuba divers. Sensors (Basel) 2017 Jun 09;17(6):1349 [FREE Full text] [doi: 10.3390/s17061349] [Medline: 28598405]
35. Eun SJ, Kim JY, Lee SH. Development of customized diving computer based on wearable sensor for marine safety. IEEE Access 2019;7:17951-17957 [FREE Full text] [doi: 10.1109/access.2019.2894740]
36. Valenko D, Mezgec Z, Pec M, Golob M. A diver's automatic buoyancy control device and its prototype development. Mar Technol Soc J 2013 Nov 01;47(6):16-26 [FREE Full text] [doi: 10.4031/mtsj.47.6.7]
37. Allotta B, Pugi L, Gelli J, Lupia M. Design and fast prototyping of a wearable safety device for divers. IFAC-PapersOnLine 2016;49(21):547-552 [FREE Full text] [doi: 10.1016/j.ifacol.2016.10.659]
38. Groh BH, Cibis T, Schill RO, Eskofier BM. IMU-based pose determination of scuba divers' bodies and shanks. 2015 Oct Presented at: 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN); June 9-12, 2015; Cambridge, MA, USA p. 1-6. [doi: 10.1109/bsn.2015.7299376]
39. Beluso C, Xu A, Patamasing E, Sebastian B, Lo E, Schurgers X, et al. D-SEA: The underwater depth sensing device for standalone time-averaged measurements. 2020 Apr Presented at: 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW); November 4-7, 2019; Monterey, CA, USA. [doi: 10.1109/massw.2019.00027]
40. Hirose M, Sugiura Y, Minamizawa K, Inami M. PukuPuCam: a recording system from third-person view in scuba diving. 2015 Mar Presented at: AH '15: The 6th Augmented Human International Conference; March 9-11, 2015; Singapore, Singapore p. 161-162. [doi: 10.1145/2735711.2735813]
41. Kuch B, Buttazzo G, Azzopardi E, Sayer M, Sieber A. GPS diving computer for underwater tracking and mapping. Underwater Technology 2012 Jul 01;30(4):189-194 [FREE Full text] [doi: 10.3723/ut.30.189]
42. McGrane O, Cronin A, Hile D. Use of handheld sonar to locate a missing diver. Wilderness Environ Med 2013 Mar;24(1):28-31. [doi: 10.1016/j.wem.2012.09.002] [Medline: 23290927]
43. Piispanen WW, Lundell RV, Tuominen LJ, Räisänen-Sokolowski AK. Assessment of alertness and cognitive performance of closed circuit rebreather divers with the critical flicker fusion frequency test in arctic diving conditions. Front Physiol 2021 Aug 10;12:722915 [FREE Full text] [doi: 10.3389/fphys.2021.722915] [Medline: 34447319]
44. Steinberg F, Doppelmayr M. Executive functions of divers are selectively impaired at 20-meter water depth. Front Psychol 2017 Jun 20;8:1000 [FREE Full text] [doi: 10.3389/fpsyg.2017.01000] [Medline: 28676772]
45. Luckerson V. Google will stop selling Glass next week. Time. 2015 Jan 15. URL: https://time.com/3669927/google-glass-explorer-program-ends/ [accessed 2022-08-08]
46. Bohn D. Intel is giving up on its smart glasses. The Verge. 2018 Apr 18. URL: https://www.theverge.com/2018/4/18/17255354/intel-vaunt-shut-down [accessed 2022-08-08]
47. Koss B, Sieber A. Development of a graphical head-up display (HUD) for rebreather diving. Underwater Technology 2011 Mar 01;29(4):203-208 [FREE Full text] [doi: 10.3723/ut.29.203]
48. Sieber A, Kuch B, Enoksson P, Stoyanova-Sieber M. Development of a head-up displayed diving computer capability for full face masks. Underwater Technology 2012 Jul 01;30(4):195-199 [FREE Full text] [doi: 10.3723/ut.30.195]
49. Gallagher DG, Manley RJ, Hughes WW, Pilcher AM. Development of an enhanced underwater navigation capability for military combat divers. 2016 Dec Presented at: OCEANS 2016 MTS/IEEE Monterey; September 19-23, 2016; Monterey, CA, USA p. 1-4. [doi: 10.1109/oceans.2016.7761159]

50. Gallagher DG, Manley RJ. Diver's full face mask head-up display system using waveguide optical display technology OSJ — 140214–001. 2015 Jan Presented at: 2014 Oceans - St. John's; September 14-19, 2014; St. John's, NL, Canada p. 1-7. [doi: 10.1109/oceans.2014.7002976]
51. Manley RJ, Gallagher DG, Hughes WW, Pilcher AMIII. Divers augmented vision display (DAVD). 2018 Jan Presented at: ASME 2017 International Mechanical Engineering Congress and Exposition; November 3–9, 2017; Tampa, Florida. [doi: 10.1115/imece2017-70026]
52. Koss B, Sieber A. Head-mounted display for diving computer platform. Journal of Display Technology 2011 Mar 11;7(4):193-199. [doi: 10.1109/jdt.2010.2103299]
53. Sieber A, Schuster A, Reif S, Madden D, Enoksson P. Head-up display system for closed circuit rebreathers with antimagnetic wireless data transmission. Mar Technol Soc J 2013 Nov 01;47(6):42-51 [FREE Full text] [doi: 10.4031/mtsj.47.6.6]
54. Cardia C, Gjanci P, Petrioli C, Saturni G, Spaccini D, Tomaselli D. The Internet of Underwater Things: From Nemo to underwater Whatsapp. 2019 Jul Presented at: Mobihoc '19: The Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing; July 2-5, 2019; Catania, Italy p. 409-410. [doi: 10.1145/3323679.3326632]
55. Anjangi P, Gibson A, Ignatius M, Pendharkar C, Kurian A, Low A, et al. Diver communication and localization system using underwater acoustics. 2020 Presented at: Global Oceans 2020: Singapore – U.S. Gulf Coast; October 5-30, 2020; Biloxi, MS p. 1-8. [doi: 10.1109/ieeeconf38699.2020.9389462]
56. Bube B, Klocke HG, Palma AML, Zanon BB. Wearable freedive computer with acoustic communication. IEEE Consumer Electronics Magazine. Epub ahead of print 2021 Jul 26. [doi: 10.1109/mce.2021.3096795]
57. Hussein K, Gonzalez JP, Derbali M. A smartphone application for two-way divers communication. 2019 Feb Presented at: 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom); March 13-15, 2019; New Delhi, India p. 1322-1325.
58. Kohlsdorf D, Gilliland S, Presti P, Starner T, Herzing D. An underwater wearable computer for two way human-dolphin communication experimentation. 2013 Sep Presented at: UbiComp '13: The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 8-12, 2013; Zurich, Switzerland p. 147-148. [doi: 10.1145/2493988.2494346]
59. Chen S, Song JL, Yuan ZM, Liu Y, Guo PP. Diver communication system based on underwater optical communication. Applied Mechanics and Materials 2014 Aug;621:259-263. [doi: 10.4028/www.scientific.net/amm.621.259]
60. Katzschmann RK, DelPreto J, MacCurdy R, Rus D. Exploration of underwater life with an acoustically controlled soft robotic fish. Sci Robot 2018 Mar 21;3(16):eaar3449. [doi: 10.1126/scirobotics.aar3449] [Medline: 33141748]
61. Kuch B, Koss B, Buttazzo G, Sieber A. Underwater navigation and communication: A novel GPS/GSM diving computer. 2009 Presented at: 35th Annual Scientific Meeting of the European Underwater and Baromedical Society (EUBS 2009); August 25-28, 2009; Aberdeen, Scotland URL: http://retis.sssup.it/~giorgio/paps/2009/eubs09-gps.pdf
62. Navea RF, Claveria MJ. Design and development of visible light communication-based underwater communication system for recreational scuba diving. International Journal of Emerging Trends in Engineering Research 2020 Jun;8(6):2266-2270 [FREE Full text] [doi: 10.30534/ijeter/2020/10862020]
63. Lee J, Jun D. Development design of wrist-mounted dive computer for marine leisure activities. Electronics (Basel) 2020 Apr 28;9(5):727 [FREE Full text] [doi: 10.3390/electronics9050727]
64. Čejka J, Chmelík J, Liarokapis F. Exploring tilting methods for typing under water. Multimed Tools Appl 2020 Aug 17;80(20):31085-31103. [doi: 10.1007/s11042-020-09305-7]
65. Wiehr F, Höh A, Krüger A. Towards a wearable for deep water blackout prevention. 2020 Mar Presented at: AHs '20: Augmented Humans International Conference; March 16-17, 2020; Kaiserslautern, Germany p. 1-3. [doi: 10.1145/3384657.3385329]
66. Renner BC, Heitmann J, Steinmetz F. ahoi: Inexpensive, low-power communication and localization for underwater sensor networks and µAUVs. ACM Trans Sens Netw 2020 May;16(2):1-46 [FREE Full text] [doi: 10.1145/3376921]
67. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: Scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
68. Singh DKA, Goh JW, Shaharudin MI, Shahar S. A mobile app (FallSA) to identify fall risk among Malaysian community-dwelling older persons: Development and validation study. JMIR Mhealth Uhealth 2021 Oct 12;9(10):e23663 [FREE Full text] [doi: 10.2196/23663] [Medline: 34636740]
69. Shawen N, Lonini L, Mummidisetty CK, Shparii I, Albert MV, Kording K, et al. Fall detection in individuals with lower limb amputations using mobile phones: Machine learning enhances robustness for real-world applications. JMIR Mhealth Uhealth 2017 Oct 11;5(10):e151 [FREE Full text] [doi: 10.2196/mhealth.8201] [Medline: 29021127]
70. Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]

Abbreviations

HUD: head-up display
IMU: inertial measurement unit
IoUT: Internet of Underwater Things
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by L Buis; submitted 07.01.22; peer-reviewed by R Chandrasekaran, D Snider, TH van de Belt; comments to author 04.04.22; revised version received 20.05.22; accepted 15.07.22; published 06.09.22.

Please cite as:
Bube B, Zanón BB, Lara Palma AM, Klocke H
Wearable Devices in Diving: Scoping Review
JMIR Mhealth Uhealth 2022;10(9):e35727
URL: https://mhealth.jmir.org/2022/9/e35727
doi:10.2196/35727
PMID:36066926

Review

mHealth Apps Using Behavior Change Techniques to Self-report Data: Systematic Review

Maria Aguiar[1,2], BSc; Maria Trujillo[2], PhD; Deisy Chaves[2,3], PhD; Roberto Álvarez[1,4], BSc; Gorka Epelde[1,4], PhD

[1]Vicomtech Foundation, Basque Research and Technology Alliance, Donostia-San Sebastián, Spain

[2]Multimedia and Computer Vision Group, Universidad del Valle, Cali, Colombia

[3]Department of Electrical, Systems and Automation, Universidad de León, León, Spain

[4]Biodonostia Health Research Institute, eHealth Group, Donostia-San Sebastián, Spain

Corresponding Author:
Maria Aguiar, BSc
Vicomtech Foundation
Basque Research and Technology Alliance
Paseo Mikeletegi 57
Donostia-San Sebastián, 20009
Spain
Phone: 34 943 30 92 30
Email: maria.aguiar@correounivalle.edu.co

Abstract

Background: The popularization of mobile health (mHealth) apps for public health or medical care purposes has transformed human life substantially, improving lifestyle behaviors and chronic condition management.

Objective: This review aimed to identify behavior change techniques (BCTs) commonly used in mHealth, assess their effectiveness based on the evidence reported in interventions and reviews to highlight the most appropriate techniques to design an optimal strategy to improve adherence to data reporting, and provide recommendations for future interventions and research.

Methods: We performed a systematic review of studies published between 2010 and 2021 in relevant scientific databases to identify and analyze mHealth interventions using BCTs that evaluated their effectiveness in terms of user adherence. Search terms included a mix of general (eg, data, information, and adherence), computer science (eg, mHealth and BCTs), and medicine (eg, personalized medicine) terms.

Results: This systematic review included 24 studies and revealed that the most frequently used BCTs in the studies were feedback and monitoring (n=20), goals and planning (n=14), associations (n=14), shaping knowledge (n=12), and personalization (n=7). However, we found mixed effectiveness of the techniques in mHealth outcomes, having more effective than ineffective outcomes in the evaluation of apps implementing techniques from the feedback and monitoring, goals and planning, associations, and personalization categories, but we could not infer causality with the results and suggest that there is still a need to improve the use of these and many common BCTs for better outcomes.

Conclusions: Personalization, associations, and goals and planning techniques were the most used BCTs in effective trials regarding adherence to mHealth apps. However, they are not necessarily the most effective since there are studies that use these techniques and do not report significant results in the proposed objectives; there is a notable overlap of BCTs within implemented app components, suggesting a need to better understand best practices for applying (a combination of) such techniques and to obtain details on the specific BCTs used in mHealth interventions. Future research should focus on studies with longer follow-up periods to determine the effectiveness of mHealth interventions on behavior change to overcome the limited evidence in the current literature, which has mostly small-sized and single-arm experiments with a short follow-up period.

(JMIR Mhealth Uhealth 2022;10(9):e33247) doi:10.2196/33247

KEYWORDS

mobile health; mHealth; behavior change techniques; adherence; app; mobile health interventions; behavior

Introduction

Overview

In modern society, the rushed lifestyle and excessive adulteration in food products have caused health-related disorders, making them an inevitable part of modern life [1]. This is associated with the development of noncommunicable diseases, which are related to 16 million premature deaths per year worldwide. The treatment of these lifestyle-related disorders demands long-term clinical help and can last a lifetime [2].

Besides, smartphones have become an essential tool in our daily lives, impacting 7.2 billion users worldwide with more than 70% of them in low- and middle-income countries [3]. Smartphone sensor technology has significantly improved and become more stable for collecting real-time data, which can be saved and processed for multiple analyses, making it possible to monitor our health through mobile health (mHealth) apps [4,5].

Recently, the popularization of mHealth apps for public health or medical care purposes have transformed human life substantially. Strategies such as reminders, counselling, reinforcement, or education have been used to improve people's adherence to the app, thus improving lifestyle behaviors [6] and chronic condition management (CCM). These strategies are known as behavior change techniques (BCTs).

There is a need to improve the adherence to one's well-being, regular health monitoring, and expert involvement [7]. The World Health Organization (WHO) estimated that in high-income countries the average adherence rate is 50% in patients with chronic medical illness [8], with even lower rates in low-income countries. It considers the extent to which a person's behavior—taking medication, following a diet, or making lifestyle changes—corresponds to recommendations agreed upon with a health professional directly or through a mobile app. Nonadherence leads to considerable morbidity, mortality, and avoidable health care costs [9], and it may be caused by people's intentional or unintentional behaviors. Intentional nonadherence refers to deciding not to report data based on the person's perceptions such as incomplete disease-related knowledge. In contrast, unintentional nonadherence means that the person intends to report data but fails because of forgetfulness or carelessness. Awareness and proper screening of these intentional and unintentional determinants for the target population are necessary to design and develop tailored solutions to ensure a methodology that improves adherence to data reporting.

mHealth has the potential to improve lifestyle and CCM, and can be rapidly adopted on a large scale and at low cost [10], but inconsistent findings have been reported on its effectiveness. Table 1 summarizes the systematic reviews conducted in the literature on the effectiveness of mHealth interventions with BCTs in distinct contexts and populations.

Although the reviews include mHealth studies for a specific population, activity, or disease, most of these studies evaluated the effectiveness in terms of the results obtained for the intervention's objective. Because of this, it is difficult to discern whether the intervention was ineffective due to a lack of adherence by participants or the combination of 35 BCTs being inadequate for the problem addressed. Therefore, the motivation of this systematic review is to identify current studies that have specifically reported their results in terms of adherence to extract the most used BCTs among the effective studies. This will help to design an adherence-focused strategy combining these BCTs.

Table 1. Systematic reviews were examined, describing the number of studies included and their research objective.

Authors	Studies, n	Objective
Schorr et al [11]	26	Identify studies using mHealth[a] for secondary CVD[b] prevention that focus on lifestyle behavior change and medication adherence
Akinosun et al [12]	25	Identify and measure the effectiveness of digital technology interventions (eg, mobile phones, the internet, software applications, or wearables) in randomized controlled trials and determine which behavior change constructs are effective at achieving risk factor modification in patients with CVD
Godinho et al [13]	29	Examine the implementation and evaluation of mHealth to support the integration of people-centered health services in the World Health Organization Western Pacific Region
Monteiro-Guerra et al [14]	17	Study real-time PA[c] coaching mobile apps with personalization mechanisms
Wang et al [15]	17	Evaluate the effectiveness of mHealth interventions for the treatment and management of diabetes and obesity reported in reviews and meta-analyses to provide recommendations for future interventions and research
Thomas Craig et al [16]	30	Identify context-aware digital behavior change interventions that provide individualized interventions to improve health
Bearne et al [17]	4	Identify apps that facilitate PA for adults with rheumatoid arthritis and compare the quality and content of these apps to incorporate relevant BCTs[d] against recommendations for cardiorespiratory, resistance, flexibility, neuromotor PA, and exercise
Kalke et al [18]	30	Identify breast cancer apps that support behavior change and assess the extent to which these apps address cancer care content
Tighe et al [19]	7	Identify digital platformlike interventions and examine their potential for supporting self-management of noncommunicable diseases and health behavior change
Armitage et al [20]	9	Estimate the efficacy of app-based interventions designed to support medication adherence and investigate which BCTs used by these apps are associated with efficacy
Pfaeffli Dale et al [21]	7	Determine the effectiveness of mHealth interventions on behavioral lifestyle changes and medication adherence for CVD self-management

[a]mHealth: mobile health.

[b]CVD: cardiovascular disease.

[c]PA: physical activity.

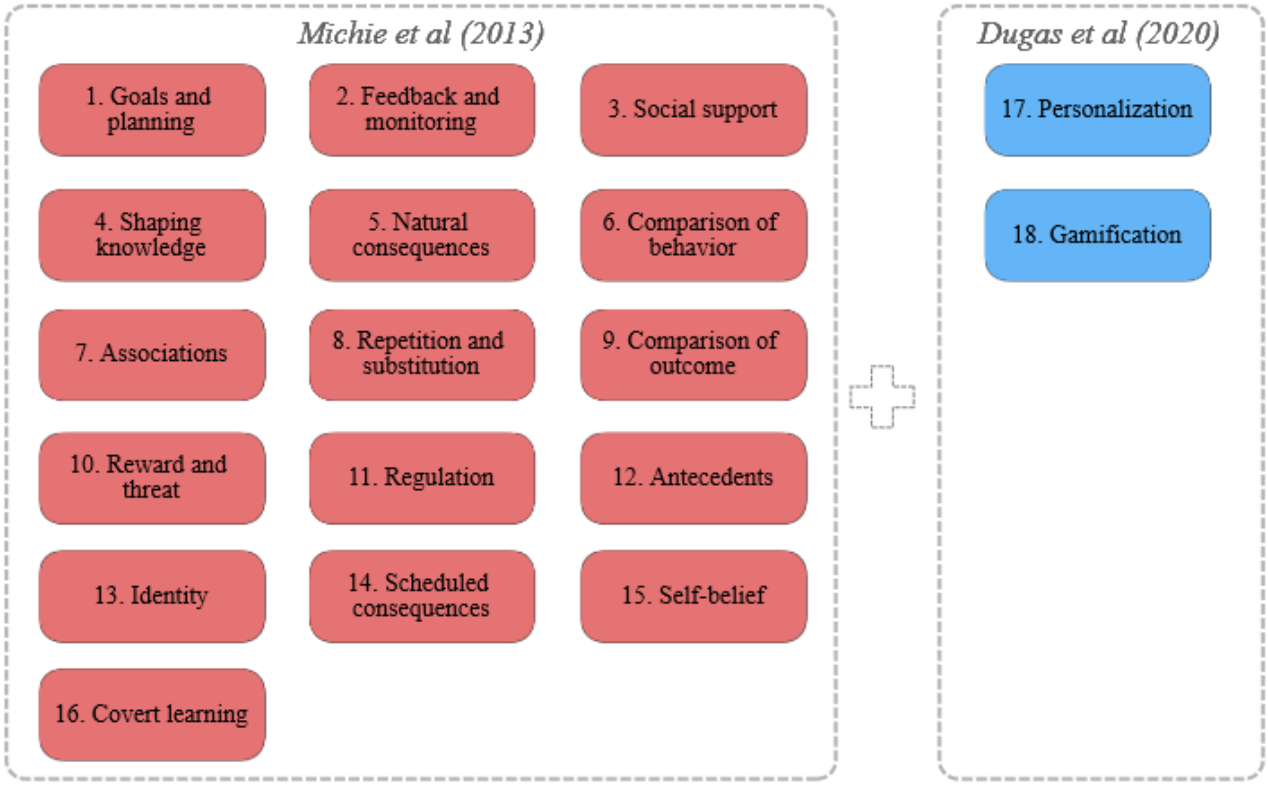
[d]BCT: behavior change technique.

Background

mHealth is a term used for mobile apps and other wearable devices that collect and monitor medical information from users [22]. Data collected from their routine leads to the accumulation of data depending on the number of users and how often they report data manually (eg, questionnaires) and through their wearable sensors. Therefore, the use of big data analytics on mHealth may be promising to provide medical information, improve people's well-being in nonclinical and clinical settings [23], and improve access to quality care and timely monitoring at an affordable cost with enriched outcomes. Within mHealth apps, BCT refers to an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior (eg, feedback, self-monitoring, and reinforcement) [24]. BCTs are coded using an established taxonomy of 93 techniques provided in "A Taxonomy of Behaviour Change Techniques Used in Interventions" [25]—for which a standardization has been proposed—and were initially grouped into 16 categories [26,27]. Besides, Dugas et al [27] extended the taxonomy of BCTs in 2020 with 2 additional categories, personalization and gamification, comprising 9 BCTs. Figure 1 shows the final taxonomy with 18 categories that will be used in this study.

Adherence is defined by the WHO [8] as "the extent to which a person's behaviour – taking medication, following a diet, and/or executing lifestyle changes – corresponds with agreed recommendations from a health care provider." Moreover, from a technical point of view, adherence is defined as the developer's expectations, referring to the degree to which the user's activity within the app matches the pattern of activity that was intended by the developers, differing from the definition of usage that refers to the level of activity within an app. For example, a user who completes 5 modules in a program will obtain 100% usage on the modules' metric of usage. However, if these modules were supposed to be completed weekly and the user only completed 3 of these on time, the user achieved 60% on the modules' adherence metric. On the other hand, if a user completes all the activities in an app by the time they are scheduled, the user adherence is 100% [7]. If people do not report data as often as expected or stop using the app, the quantity and homogeneity of data to be processed will be reduced, producing lower quality outcomes. We will consider both definitions when reviewing the efficacy of the studies.

Figure 1. Taxonomy: behavior change techniques categories.



Methods

Research Questions

The objective of our systematic review is to identify and analyze relevant studies on BCTs used in mHealth interventions, focusing on the effectiveness of the BCTs on use and paying special attention to improve adherence to data reporting. Hence, the research questions (RQs) that have guided this review are:

- RQ1: What BCTs are most commonly used in the context of mHealth apps? How can these techniques be classified?
- RQ2: What is the performance of these techniques concerning end-user adherence?
- RQ3: Are BCTs (personalization, feedback, and monitoring strategies) useful for improving adherence to data reporting in mHealth apps?

Search Strategy

This systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [28]. Details of the search strategy are given below.

Eligibility Criteria

Eligible studies were peer-reviewed articles published in English from January 1, 2010, to December 2021. With the growth of technology used for mHealth, the selected time allowed us to assess studies using the most relevant technologies. Participants in the studies were patients of any age, both healthy and with any type of chronic disease. Moreover, studies with interventions using health practices supported by any type of mobile device were eligible for inclusion. Besides, interventions could include multiple delivery methods and nondigital elements. The result assessed was adherence focusing on any outcome (eg, physical activity, medication adherence, and data reporting) measured by any metric during any follow-up period.

Selection of Sources

The scientific databases selected for the review were Scopus, PubMed, Web of Science, and IEEE Xplore. The combination of these databases provides comprehensive coverage of publications in the context of medical informatics, high relevance, and a complete advanced search. Search strings and search methods were consistent across all databases.

Search Terms

A preliminary literature review resulted in the first search equation, which aimed to find the different applications of mHealth in personalized medicine. This allowed us to identify relevant keywords and search terms to refine the search equation in each iteration by focusing on relevant topics such as adherence and data reporting but without limiting it to the field of personalized medicine. Finally, behavior change was found to be a popular aspect of improving adherence and hence the effectiveness of studies. We only included the term “behaviour change” because it is not limited to techniques, as there are different theories and models that contain them. As a result, four search iterations were performed. Those iterations are shown in Table 2.

Table 2. Search equations, data source, and total records per query.

Search equation and source	Total records, n
1 ("mHealth" AND "personalised medicine")	
Scopus	124
PubMed	19
Web of Science	4
IEEE Xplore	15
2 ("mobile health" OR "mHealth") AND "adherence" AND "personalised medicine")	
Scopus	23
PubMed	2
Web of Science	0
IEEE Xplore	5
3 ("mHealth" AND ("data" OR "information") AND "registration" AND "adherence")	
Scopus	52
PubMed	152
Web of Science	35
IEEE Xplore	1
4 ("mobile health" OR "mHealth") AND "adherence" AND "behaviour change")	
Scopus	156
PubMed	120
Web of Science	23
IEEE Xplore	9

Inclusion/Exclusion Criteria

After the initial gathering and screening of studies, articles were selected based on predefined eligibility criteria.

Inclusion Criteria

The included studies meet the criteria of being an mHealth intervention. In addition, they fulfill at least one of the following conditions:

- The study implements at least one BCT.
- The study compares several BCTs.
- The study evaluates the BCTs using at least one adherence metric.

Exclusion Criteria

The excluded studies meet at least one of the following conditions that aim to discard irrelevant studies for this systematic review:

- The study does not address the use of BCTs in the context of mHealth or vice versa.
- The study does not present at least one of its results in terms of adherence.
- The study was not published in between the years 2010-2021.
- The study does not belong to one of these categories: journal paper, conference paper, or review.
- The study was not peer reviewed.
- The study is not written in English or Spanish.

Data Extraction and Coding of BCTs

PRISMA [28] guidelines were used for data extraction. We gathered information about the study background (year, authors, etc), eligibility criteria (population), the number of participants, intervention description, technology, and results.

The behavioral strategy used in each study was identified and coded using the taxonomy of 18 categories introduced in the Background section to answer RQ1. To perform correct identification and coding, free taxonomy training was received using materials available online [29]. Results using any adherence assessment metric were extracted to solve RQ2. Finally, to answer RQ3, a more in-depth analysis was performed in the Discussion section about what was found to answer the previous RQs.

Results

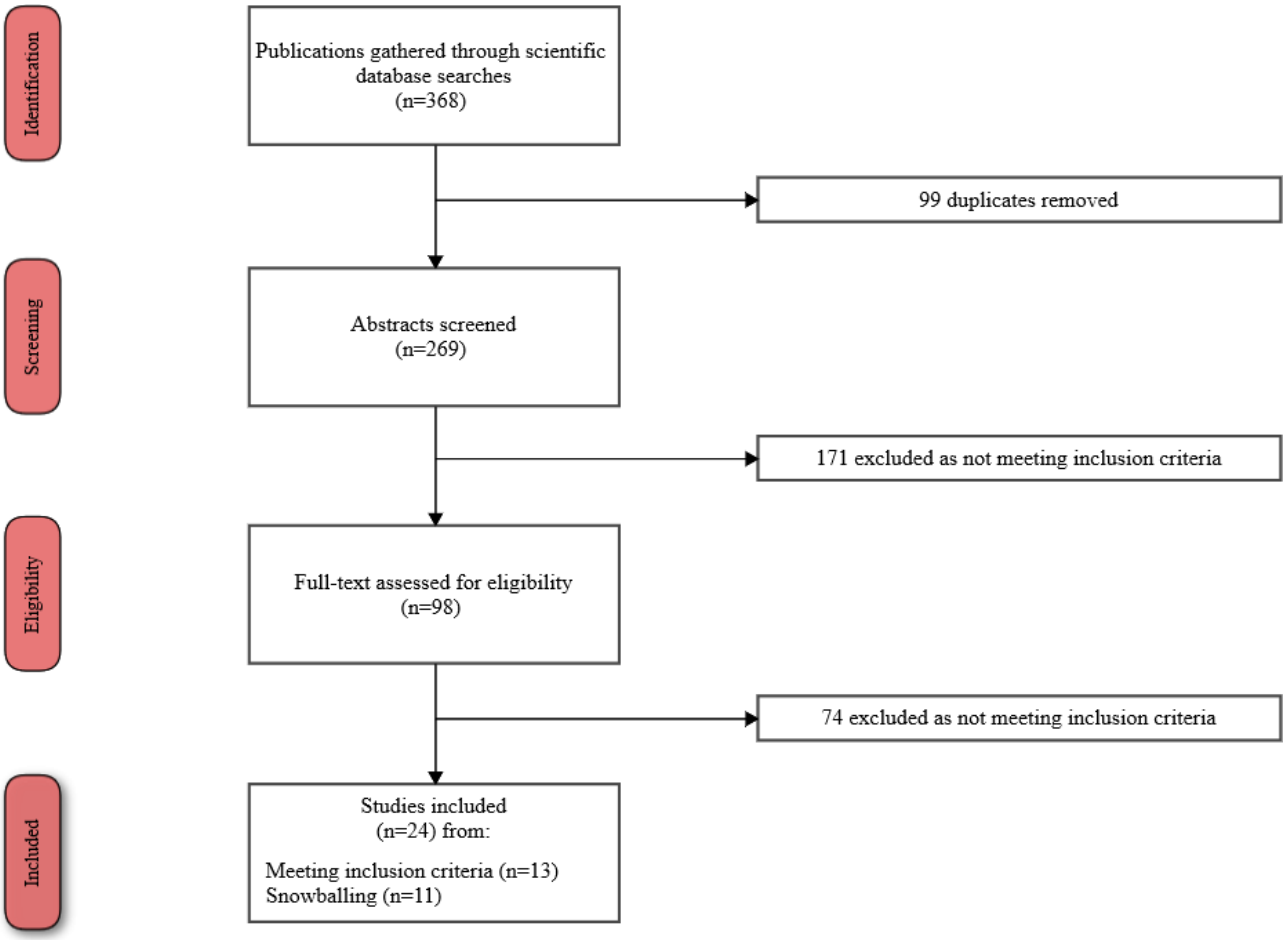
This section presents the results obtained from the methodology described in the Methods. First, an overview of the selected studies and their main characteristics is presented, and then an analysis of these studies is performed.

Search Results and Study Selection

The search results and study selection are summarized in the PRISMA flowchart in Figure 2. A total of 368 studies were obtained from the search engines. Subsequently, 88 duplicate studies obtained from the combination of the databases' results were excluded using the Mendeley tool. In the screening step,

we had 269 studies, of which 171 were excluded for not meeting the inclusion criteria after abstract review. In the eligibility step, we had 98 studies, of which 74 were excluded for not meeting the inclusion criteria after full-text review. Finally, 13 studies were included; however, the snowball strategy was applied to the systematic reviews listed in Table 1, resulting in 11 additional studies. These studies were not found in our initial search because they did not use the term adherence in their title, abstract, or keywords.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Outcome and Characteristics

We summarized the selected 24 studies [30-53] to understand which technologies and BCTs were most frequently used in mHealth research. Data extracted from each study following the methodology defined in the Data Extraction and Coding of BCTs section are presented in Table S1 in Multimedia Appendix 1 [30-53]. The first column refers to the study identifier, the next two columns specify the sample, and the remaining four columns correspond to the duration, description, technology, and evaluation of the intervention.

We observed in the reviewed papers that the most targeted behavior was medication adherence—present in 67% (n=16) of the studies—which is important in preventing rehospitalization, morbidity, mortality, and increased health care costs [54]. We also found that the most frequent population was patients diagnosed with chronic conditions, participating in 83% (n=20) of the studies, and the most popular form of technology in the studies was the use of apps, present in 71% (n=17) of the studies, with wearable support becoming popular. Still, 29% (n=7) of the studies examined the effects of SMS text message–based interventions, indicating that simpler health interventions delivered by text remain popular. With respect to the techniques, the most frequently used BCTs in the studies were feedback and monitoring (n=20, 83%), goals and planning (n=14, 58%), associations (n=14, 58%), and shaping knowledge (n=12, 50%). On the other hand, personalization (n=7, 29%) was approached in a simple way, considering their capabilities, including tailoring to demographic information, health status (eg, alcoholic or nonalcoholic), and time of notifications. On average, 4 BCTs are included in each mHealth intervention; a minimum of 1 and maximum of 8 techniques were used per study. This is shown in Table S2 in Multimedia Appendix 2 [30-53] where the BCTs used by each study are presented.

Regarding adherence evaluation metrics, it is worth mentioning that different methods are used for this measurement, from daily use of the mobile app to the Morisky Medication Adherence Scale. To analyze these studies homogeneously, a cutoff point for effectiveness in terms of adherence rate was considered. High adherence was defined by an adherence rate ≥80% and nonadherent as an adherence rate <80%. This cut point is conventional in the adherence literature [55,56] and is considered crucial for the effectiveness of long-term therapy [57]; however, it is interesting to note that many studies used

the arbitrary threshold of 80% [58], indicating that the optimal cut point for adherence ranged from 58% to 85%.

Moreover, after the BCTs used in each study were identified, we classified the studies into effective and ineffective using the cut point of 80% to gain further insight about why some interventions yielded significant improvements. As a result, Table 3 shows the number of times each BCT was used in both effective and ineffective studies.

This comparison revealed that, among the top five most used BCTs in the studies, feedback and monitoring (60%-40%) along with associations (57%-43%) have been used homogeneously in effective and ineffective studies, goals and planning (79%-21%) and personalization (71%-29%) have improved the effectiveness of the interventions since they have a higher presence in effective than in ineffective studies, and shaping knowledge (42%-58%) had a lower presence in effective than ineffective studies.

It was observed that 2 of the main BCTs (goals and planning, and associations) used on effective interventions were present in 100% of the adherence-oriented studies for mobile apps and in none of the ineffective studies [30,40,42,52]. Meanwhile, shaping knowledge was present in 100% of the ineffective studies for medication adherence and in none of the effective ones [31,33,35,37,44,47]. This is a good starting point for the design of a general BCT strategy.

Table 3. Behavior change techniques used in effective and ineffective studies (N=24).

Behavior change technique	Effective studies (≥80% adherence rate), n (%)	Ineffective studies (<80% adherence rate), n (%)
Feedback and monitoring	12 (50)	8 (33)
Goals and planning	11 (46)	3 (13)
Associations	8 (33)	6 (25)
Shaping knowledge	5 (21)	7 (29)
Personalization	5 (21)	2 (8)
Regulation	4 (17)	3 (13)
Reward and threat	4 (17)	1 (4)
Social support	4 (17)	1 (4)
Comparison of behavior	2 (8)	3 (13)
Natural consequences	3 (13)	0 (0)
Repetition and substitution	2 (8)	0 (0)
Antecedents	1 (4)	1 (4)
Comparison of outcomes	0 (0)	1 (4)

Discussion

Principal Findings

We answered three RQs (defined in the Methods section) related to BCTs used in mHealth interventions. To answer RQ1, the taxonomy of BCTs, proposed by Michie et al [24] and which has been updated over the years [27], was identified. Based on this, the BCTs used in the selected studies were extracted and coded to the 18 categories to improve the understanding and meaning of the comparisons. We found that the most frequently used BCTs in mHealth interventions were *feedback and monitoring*, *goals and planning*, *associations*, *shaping knowledge*, and *personalization*. Similar BCTs were also found to be common in reviews of mHealth interventions targeting physical activity and sedentary behaviors, lifestyle, and medication adherence for CCM [59,60].

For RQ2, the categories of BCTs with the highest presence in effective studies were goals and planning, associations, feedback and monitoring, and personalization (as described in the Results section). However, in line with other reviews that found mixed effectiveness in mHealth outcomes [15,27], they suggested that the need remains to improve the use of these and many common BCTs for better outcomes. Possible reasons for the mixed effects of BCTs include the need for details on the specific BCTs that are used in the studies since current mHealth interventions often lump together a multitude of BCTs, making it difficult to discern the characteristics that lead to a study being effective or not and for more specific mHealth intervention content for different population subgroups (eg, those with specific mental health disorders such as anxiety) who may react differently compared to those with healthy mental health in an intervention setting, such as patients with coronary heart disease. This was reflected in the finding that some BCTs did not have the same effects for all groups, with those with higher levels of depression or anxiety deriving less benefit from some BCTs.

Regarding RQ3, considering the responses for RQ1 and RQ2, it was observed that, although some studies in the mHealth context have evaluated user adherence as a complementary outcome resulting from their implementation of BCTs, none of the reviewed studies applied BCTs for adherence to data reporting or use of the mHealth app; the studies generally aimed to improve outcomes for activity in chronically ill or healthy people. However, designing an approach that combines the BCTs most effective for adherence (feedback and monitoring, goals and planning, associations, and personalization), previously identified in RQ2, could be helpful to improve data reporting in mHealth apps.

Limitations

Several limitations were found that are important to highlight. This review focused on mHealth interventions that include BCTs and how they affect user adherence. As a result, we found a lack of heterogeneity in the sample of results, evidencing that most of these interventions do not focus on improving user adherence to data reporting or to the app but rather on achieving an improvement in the objective of the study, whether it is an improvement in lifestyle behaviors or CCM.

In addition, designing mHealth apps is relatively new, and there is little agreement regarding best practices. Researchers are still trying to understand how and why BCTs lead to positive health outcomes when delivered through apps. There is growing evidence to support the effectiveness of mHealth interventions on health outcomes. Even though the evidence is growing, it is still relatively weak. This may be in part due to the difficulty of designing and conducting rigorous studies on mHealth interventions, even more in the context of the COVID-19 pandemic.

Additionally, the BCT taxonomy approach used to summarize the characteristics of interventions still has some limitations, despite the inclusion of the extension proposed by Dugas et al [27]. This taxonomy allows for the coding of different BCTs but does not assess the intensity or dosage of interventions. Nevertheless, the taxonomy works perfectly well as an excellent starting point and a standard to systematically describe an mHealth intervention.

Finally, we found that most studies are small single-arm studies with short follow-up periods. Future research should focus on studies with longer follow-up periods to determine the effectiveness of mHealth interventions on behavior change.

Conclusions

mHealth interventions to improve lifestyle behaviors and CCM have become popular in recent years, improving along with different technologies. Although SMS text message–based mHealth interventions remained popular for their proven high results in terms of effectiveness, this review suggests that mHealth is progressively moving toward the implementation of mobile apps and wearable interventions, replacing SMS text messaging with notifications. This review also revealed that mHealth is being applied to address a diversity of lifestyle behaviors and health outcomes, showing its applicability to a variety of health care contexts, with one of its focuses being medication adherence. On the other hand, we found that the most frequently used BCTs in mHealth interventions were feedback and monitoring, goals and planning, associations, shaping knowledge, and personalization. However, this does not necessarily imply that they are the most effective, so we conducted further analysis that found that frequently used BCTs in ineffective studies are often well supported in the health behavior change literature [61,62], suggesting a need to better understand best practices for applying such techniques and to obtain details on the specific BCTs used in mHealth interventions.

Congruent with other reviews that found heterogeneous effectiveness of mHealth, the results also suggested that there remains a need to optimize the use of BCTs or find a better combination for better outcomes, it is intended that this review could help identify the most appropriate techniques to improve adherence to data reporting and thus design an optimal strategy while taking into account the differences among population subgroups, pointing to the need to go beyond the idea that “one-size-fits-all.” As a small step forward, more sophisticated technologies such as wearable activity trackers and wireless sensors have been included in mHealth interventions.

Acknowledgments

This research was developed within the BG19 (2019-2020/Grant number KK-2019/00032) and the BG21 (2021-2022/Grant number KK-2021/00005) research projects funded by the Department of Economic Development and Infrastructure of the Basque Government under the Elkartek program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile health studies describing the sample criteria (population), the sample size (N), the study duration (duration), the technology used (tech), and results.
[DOC File , 38 KB - mhealth_v10i9e33247_app1.doc]

Multimedia Appendix 2
Behavior change techniques in mobile health studies.
[DOC File , 34 KB - mhealth_v10i9e33247_app2.doc]

References

1. Saxena M, Saxena A. Evolution of mHealth eco-system: a step towards personalized medicine. In: Khanna A, Gupta D, Bhattacharyya S, Snasel V, Platos J, Hassanien AE, editors. International Conference on Innovative Computing and Communications Proceedings of ICICC 2019, Volume 1. Singapore: Springer; 2020:351-379.
2. Lapão LV. The future of healthcare: the impact of digitalization on healthcare services performance. In: Neto AP, Flynn MB, editors. The Internet and Health in Brazil: Challenges and Trends. Cham: Springer; 2019:435-449.
3. Paglialonga A, Mastropietro A, Scalco E, Rizzo G. The mHealth. In: Andreoni G, Perego P, Frumento E, editors. m_Health Current and Future Applications. Cham: Springer; 2019:5-17.
4. Weiler A. mHealth and big data will bring meaning and value to patient-reported outcomes. Mhealth 2016;2:2. [doi: 10.3978/j.issn.2306-9740.2016.01.02] [Medline: 28293580]
5. Kubben P. Mobile apps. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer; 2019:171-179.
6. Hsieh W, Su Y, Han H, Huang M. A novel mHealth approach for a patient-centered medication and health management system in Taiwan: pilot study. JMIR Mhealth Uhealth 2018 Jul 03;6(7):e154 [FREE Full text] [doi: 10.2196/mhealth.9987] [Medline: 29970356]
7. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
8. Sabate E. Adherence to long-term therapies: evidence for action. World Health Organization. 2003 May 28. URL: https://apps.who.int/iris/handle/10665/42682?locale-attribute=es& [accessed 2022-09-01]
9. McDonald HP, Garg AX, Haynes RB. Interventions to enhance patient adherence to medication prescriptions: scientific review. JAMA 2002 Dec 11;288(22):2868-2879. [doi: 10.1001/jama.288.22.2868] [Medline: 12472329]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Schorr EN, Gepner AD, Dolansky MA, Forman DE, Park LG, Petersen KS, Thrombosis and Vascular Biology; and Council on Lifestyle and Cardiometabolic Health. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103 [FREE Full text] [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]
12. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Mar 03;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]
13. Godinho MA, Jonnagaddala J, Gudi N, Islam R, Narasimhan P, Liaw S. mHealth for integrated people-centred health services in the Western Pacific: a systematic review. Int J Med Inform 2020 Oct;142:104259. [doi: 10.1016/j.ijmedinf.2020.104259] [Medline: 32858339]
14. Monteiro-Guerra FM, Rivera-Romero O, Fernandez-Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2020 Jun;24(6):1738-1751. [doi: 10.1109/JBHI.2019.2947243] [Medline: 31751254]
15. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
16. Thomas Craig KJ, Morgan LC, Chen C, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048 [FREE Full text] [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
17. Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e18495 [FREE Full text] [doi: 10.2196/18495] [Medline: 32706727]
18. Kalke K, Ginossar T, Bentley JM, Carver H, Shah SFA, Kinney AY. Use of evidence-based best practices and behavior change techniques in breast cancer apps: systematic analysis. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e14082 [FREE Full text] [doi: 10.2196/14082] [Medline: 32012084]
19. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
20. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045 [FREE Full text] [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]
21. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. Eur J Prev Cardiol 2016 May;23(8):801-817. [doi: 10.1177/2047487315613462] [Medline: 26490093]

22. Perego P. Device for mHealth. In: Andreoni G, Perego P, Frumento E, editors. m_Health Current and Future Applications. Cham: Springer; 2019:87-99.
23. van den Berg A, Mummery CL, Passier R, van der Meer AD. Personalised organs-on-chips: functional testing for precision medicine. Lab Chip 2019 Jan 15;19(2):198-205 [FREE Full text] [doi: 10.1039/c8lc00827b] [Medline: 30506070]
24. 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]
25. 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]
26. BCT Taxonomy (v1): 93 hierarchically-clustered techniques. Digitalwellbeing.org. URL: https://digitalwellbeing.org/wp-content/uploads/2016/11/BCTTv1_PDF_version.pdf [accessed 2021-08-05]
27. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]
28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
29. BCT Taxonomy. URL: http://www.bct-taxonomy.com/ [accessed 2021-08-05]
30. Klasnja P, Rosenberg DE, Zhou J, Anau J, Gupta A, Arterburn DE. A quality-improvement optimization pilot of BariFit, a mobile health intervention to promote physical activity after bariatric surgery. Transl Behav Med 2021 Mar 16;11(2):530-539. [doi: 10.1093/tbm/ibaa040] [Medline: 32421187]
31. Jahan Y, Rahman MM, Faruque ASG, Chisti MJ, Kazawa K, Matsuyama R, et al. Awareness development and usage of mobile health technology among individuals with hypertension in a rural community of Bangladesh: randomized controlled trial. J Med Internet Res 2020 Dec 07;22(12):e19137 [FREE Full text] [doi: 10.2196/19137] [Medline: 33284129]
32. Eaton C, Comer M, Pruette C, Psoter K, Riekert K. Text messaging adherence intervention for adolescents and young adults with chronic kidney disease: pilot randomized controlled trial and stakeholder interviews. J Med Internet Res 2020 Aug 14;22(8):e19861 [FREE Full text] [doi: 10.2196/19861] [Medline: 32795983]
33. Mata J, Pecorelli N, Kaneva P, Moldoveanu D, Gosselin-Tardiff A, Alhashemi M, et al. A mobile device application (app) to improve adherence to an enhanced recovery program for colorectal surgery: a randomized controlled trial. Surg Endosc 2020 Feb;34(2):742-751. [doi: 10.1007/s00464-019-06823-w] [Medline: 31087175]
34. Janevic MR, Shute V, Murphy SL, Piette JD. Acceptability and effects of commercially available activity trackers for chronic pain management among older African American adults. Pain Med 2020 Feb 01;21(2):e68-e78 [FREE Full text] [doi: 10.1093/pm/pnz215] [Medline: 31509196]
35. Fico G, Martinez-Millana A, Leuteritz J, Fioravanti A, Beltrán-Jaunsarás ME, Traver V, et al. User centered design to improve information exchange in diabetes care through eHealth : results from a small scale exploratory study. J Med Syst 2019 Nov 18;44(1):2. [doi: 10.1007/s10916-019-1472-5] [Medline: 31741069]
36. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 06;16(7):1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
37. Coorey G, Peiris D, Usherwood T, Neubeck L, Mulley J, Redfern J. Persuasive design features within a consumer-focused eHealth intervention integrated with the electronic health record: a mixed methods study of effectiveness and acceptability. PLoS One 2019;14(6):e0218447 [FREE Full text] [doi: 10.1371/journal.pone.0218447] [Medline: 31220127]
38. Hovland-Tånneryd A, Melin M, Hägglund E, Hagerman I, Persson HE. From randomised controlled trial to real world implementation of a novel home-based heart failure tool: pooled and comparative analyses of two clinical controlled trials. Open Heart 2019;6(1):e000954 [FREE Full text] [doi: 10.1136/openhrt-2018-000954] [Medline: 31217992]
39. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP Randomized Clinical Trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
40. Svendsen MT, Andersen F, Andersen KH, Pottegård A, Johannessen H, Möller S, et al. A smartphone application supporting patients with psoriasis improves adherence to topical treatment: a randomized controlled trial. Br J Dermatol 2018 Nov;179(5):1062-1071. [doi: 10.1111/bjd.16667] [Medline: 29654699]
41. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
42. Lakshminarayana R, Wang D, Burn D, Chaudhuri KR, Galtrey C, Guzman NV, et al. Using a smartphone-based self-management platform to support medication adherence and clinical consultation in Parkinson's disease. NPJ Parkinsons Dis 2017;3:2. [doi: 10.1038/s41531-016-0003-z] [Medline: 28649602]
43. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation

via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446. [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
44. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MDC, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
45. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
46. Ammenwerth E, Woess S, Baumgartner C, Fetz B, van der Heidt A, Kastner P, et al. Evaluation of an integrated telemonitoring surveillance system in patients with coronary heart disease. Methods Inf Med 2015;54(5):388-397. [doi: 10.3414/ME15-02-0002] [Medline: 26395147]
47. Safran Naimark J, Madar Z, Shahar DR. The impact of a web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. J Med Internet Res 2015 Mar 02;17(3):e56 [FREE Full text] [doi: 10.2196/jmir.3682] [Medline: 25732936]
48. Hammonds T, Rickert K, Goldstein C, Gathright E, Gilmore S, Derflinger B, et al. Adherence to antidepressant medications: a randomized controlled trial of medication reminding in college students. J Am Coll Health 2015;63(3):204-208. [doi: 10.1080/07448481.2014.975716] [Medline: 25338175]
49. Goldstein CM, Gathright EC, Dolansky MA, Gunstad J, Sterns A, Redle JD, et al. Randomized controlled feasibility trial of two telemedicine medication reminder systems for older adults with heart failure. J Telemed Telecare 2014 Sep;20(6):293-299 [FREE Full text] [doi: 10.1177/1357633X14541039] [Medline: 24958355]
50. Santo K, Singleton A, Rogers K, Thiagalingam A, Chalmers J, Chow CK, et al. Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial. Heart 2019 Feb;105(4):323-329. [doi: 10.1136/heartjnl-2018-313479] [Medline: 30150326]
51. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
52. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 05;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
53. Párraga-Martínez I, Rabanales-Sotos J, Lago-Deibe F, Téllez-Lapeira JM, Escobar-Rabadán F, Villena-Ferrer A, et al. Effectiveness of a combined strategy to improve therapeutic compliance and degree of control among patients with hypercholesterolaemia: a randomised clinical trial. BMC Cardiovasc Disord 2015 Jan 19;15:8 [FREE Full text] [doi: 10.1186/1471-2261-15-8] [Medline: 25599690]
54. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
55. Wei L, Wang J, Thompson P, Wong S, Struthers AD, MacDonald TM. Adherence to statin treatment and readmission of patients after myocardial infarction: a six year follow up study. Heart 2002 Sep;88(3):229-233 [FREE Full text] [doi: 10.1136/heart.88.3.229] [Medline: 12181210]
56. Chapman RH, Benner JS, Petrilla AA, Tierce JC, Collins SR, Battleman DS, et al. Predictors of adherence with antihypertensive and lipid-lowering therapy. Arch Intern Med 2005 May 23;165(10):1147-1152. [doi: 10.1001/archinte.165.10.1147] [Medline: 15911728]
57. Bansilal S, Castellano JM, Garrido E, Wei HG, Freeman A, Spettell C, et al. Assessing the impact of medication adherence on long-term cardiovascular outcomes. J Am Coll Cardiol 2016 Aug 23;68(8):789-801 [FREE Full text] [doi: 10.1016/j.jacc.2016.06.005] [Medline: 27539170]
58. Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. Curr Med Res Opin 2009 Sep;25(9):2303-2310. [doi: 10.1185/03007990903126833] [Medline: 19635045]
59. 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]
60. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
61. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
62. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

Abbreviations

BCT: behavior change technique
CCM: chronic condition management
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
WHO: World Health Organization

Edited by L Buis; submitted 09.09.21; peer-reviewed by L Park, J Lauffenburger, A Fioravanti; comments to author 01.11.21; revised version received 15.03.22; accepted 03.08.22; published 09.09.22.

Please cite as:
Aguiar M, Trujillo M, Chaves D, Álvarez R, Epelde G
mHealth Apps Using Behavior Change Techniques to Self-report Data: Systematic Review
JMIR Mhealth Uhealth 2022;10(9):e33247
URL: https://mhealth.jmir.org/2022/9/e33247
doi:10.2196/33247
PMID:36083606

Review

Wearables for Measuring Health Effects of Climate Change–Induced Weather Extremes: Scoping Review

Mara Koch[1]; Ina Matzke[1]; Sophie Huhn[1]; Hanns-Christian Gunga[2], Prof Dr; Martina Anna Maggioni[2,3], PD, PhD; Stephen Munga[4], PhD; David Obor[4], MSc; Ali Sié[1,5], PhD; Valentin Boudo[5], MSc; Aditi Bunker[1,6], Dr sc hum; Peter Dambach[1], PD, PhD; Till Bärnighausen[1,7,8], Prof Dr; Sandra Barteit[1], Dr sc hum

[1]Heidelberg Institute of Global Health, Faculty of Medicine and University Hospital, Heidelberg University, Heidelberg, Germany

[2]Charité - Universitätsmedizin Berlin, Institute of Physiology, Center for Space Medicine and Extreme Environment Berlin, Berlin, Germany

[3]Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy

[4]Kenya Medical Research Institute, Kisumu, Kenya

[5]Centre de Recherche en Santé, Nouna, Burkina Faso

[6]Center for Climate, Health, and the Global Environment, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[7]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[8]Africa Health Research Institute, KwaZulu-Natal, South Africa

Corresponding Author:
Mara Koch
Heidelberg Institute of Global Health
Faculty of Medicine and University Hospital
Heidelberg University
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 6221 5634030
Email: mara.koch@uni-heidelberg.de

Abstract

Background: Although climate change is one of the biggest global health threats, individual-level and short-term data on direct exposure and health impacts are still scarce. Wearable electronic devices (wearables) present a potential solution to this research gap. Wearables have become widely accepted in various areas of health research for ecological momentary assessment, and some studies have used wearables in the field of climate change and health. However, these studies vary in study design, demographics, and outcome variables, and existing research has not been mapped.

Objective: In this review, we aimed to map existing research on wearables used to detect direct health impacts and individual exposure during climate change–induced weather extremes, such as heat waves or wildfires.

Methods: We conducted a scoping review according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) framework and systematically searched 6 databases (PubMed [MEDLINE], IEEE Xplore, CINAHL [EBSCOhost], WoS, Scopus, Ovid [MEDLINE], and Google Scholar). The search yielded 1871 results. Abstracts and full texts were screened by 2 reviewers (MK and IM) independently using the inclusion and exclusion criteria. The inclusion criteria comprised studies published since 2010 that used off-the-shelf wearables that were neither invasive nor obtrusive to the user in the setting of climate change–related weather extremes. Data were charted using a structured form, and the study outcomes were narratively synthesized.

Results: The review included 55,284 study participants using wearables in 53 studies. Most studies were conducted in upper–middle-income and high-income countries (50/53, 94%) in urban environments (25/53, 47%) or in a climatic chamber (19/53, 36%) and assessed the health effects of heat exposure (52/53, 98%). The majority reported adverse health effects of heat exposure on sleep, physical activity, and heart rate. The remaining studies assessed occupational heat stress or compared individual- and area-level heat exposure. In total, 26% (14/53) of studies determined that all examined wearables were valid and reliable for measuring health parameters during heat exposure when compared with standard methods.

Conclusions: Wearables have been used successfully in large-scale research to measure the health implications of climate change–related weather extremes. More research is needed in low-income countries and vulnerable populations with pre-existing

conditions. In addition, further research could focus on the health impacts of other climate change–related conditions and the effectiveness of adaptation measures at the individual level to such weather extremes.

(JMIR Mhealth Uhealth 2022;10(9):e39532) doi:10.2196/39532

KEYWORDS

wearable; consumer-grade wearables; fitness trackers; climate change; heat; global health; public health; review; mobile phone

Introduction

Background

Climate change is one of the biggest global health threats of the century [1], and the field of climate and health research has been rapidly growing [2]. Many environmental conditions such as rising temperatures, floods, wildfires, heat waves, droughts, and other extreme weather events can be linked to climate change according to the 2021 Intergovernmental Panel on Climate Change report [3] and may, directly and indirectly, impact human health [4]. The wide-ranging health effects of these weather extremes include malnutrition from food insecurity; infectious disease; respiratory, cardiovascular, neurological, and mental health disorders; and mortality [4,5].

Epidemiological studies often focus on the relationship between heat and mortality or morbidity in terms of the number of hospital admissions or long-term effects but do not consider individual exposure and direct health effects [5]. Furthermore, most studies use weather and climate data from satellites or the nearest weather station, which is often located at the airport. These approaches do not consider granular spatial and temporal differences in weather exposure or individual factors that influence the exposure such as time spent indoors [6,7]. To this end, consumer-grade wearable devices (hereafter *wearables*) could generate high-resolution data at the individual level, measuring exposure and health parameters in the real-life environment, the ecological momentary assessment [8]. Wearables can cover a variety of variables and physiological data, including, among others, activity levels, sleep, sweat rate, and heart rate (HR) [9], presenting a potential solution to the shortage of short-term and individual-level data in climate change and health research.

In recent years, some reviews have been conducted on the assessment of heat strain and individual heat exposure using wearable devices. However, these studies have mainly focused on urban and occupational heat exposure [10,11], although populations living in low- and middle-income countries and rural settings have a high vulnerability to climate change [12]. Although the urban heat island effect describes higher heat exposure in cities owing to human activities and dense concentrations of surfaces that absorb and retain heat, rural populations are often more exposed because of their reliance on climate-sensitive livelihoods [10,12]. Some reviews have examined the validity of various wearables but only in moderate climate settings [13,14]. Furthermore, many studies [15,16] used prototypes and not off-the-shelf devices, which make them difficult to reproduce in the field.

Research Objectives

Therefore, the overarching objectives of this review were (1) to map the available research on the use of off-the-shelf wearables for measuring direct health effects of and individual exposure to climate change–induced weather extremes such as heat, (2) to examine current approaches to wearable use in this field, and (3) to identify gaps in the research. We particularly focused on (1) demographic characteristics, (2) selected wearable devices and their measures, (3) extreme weather condition exposure and data collection methods, (4) analytical approaches, (5) validity of wearables in extreme weather conditions, and (6) observed effects of extreme weather exposure on health (especially of heat on sleep, physical activity, and HR, as well as occupational heat stress).

Methods

Overview

The methodology for this scoping review was based on the framework outlined by Arksey and O'Malley [17] and Peters et al [18] and in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [19] (Multimedia Appendix 1). A review protocol can be obtained from the principal author (MK) upon request. A scoping review seemed most appropriate to approach the research objective, as initial research into this topic revealed a broad scope of heterogeneous studies, however, limited in their numbers.

Eligibility Criteria

We included articles that were available in English and published after January 1, 2010, because wearables have become widely available on the consumer market and were also increasingly adopted in research since then [20,21]. Types of studies included were case studies, observational studies, non–randomized controlled trials, and randomized controlled trials. We included any consumer- or research-grade wearables that were available off-the-shelf, could be worn on the body, and were neither invasive nor obtrusive (excluding, eg, ingestible, handheld, or wired devices). All types of sensors or measurements that measured at least one physiological parameter or individual exposure were included. For a complete list of eligibility criteria, see Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Publications:
  - Full text available
  - Published in the English language
  - Published between January 2010 and September 2021
  - Randomized controlled trials (RCTs), non-RCTs, observational studies, or case studies
- Wearable device:
  - Off-the-shelf wearable electronic devices
  - Noninvasive and nonobtrusive
  - Measuring at least one physiological parameter (eg, heart rate or sleep duration) or individual exposure (eg, ambient temperature)
- Climate change:
  - Climate change–related weather extremes: heat, flood, drought, wildfire, tropical cyclone, or heavy precipitation
  - Exposure: outdoors, indoors, or in a climatic chamber or laboratory
- Outcomes:
  - Individual effect of climate change–related environmental condition measured with wearables
  - Validity and method comparison of wearables in extreme weather conditions

Exclusion criteria

- Publications:
  - Nonhuman study population
  - Reviews, editorials, or commentaries
- Wearable device:
  - Not commercially available (eg, prototype or design study)
  - Wearable with interventional function only (eg, cooling vest)
  - Smartphone used as a wearable
  - Wearable not implemented
- Climate change:
  - Other environmental conditions
  - Exposure to heat in the context of mining or firefighting
- Outcomes:
  - Wearable (data) not specifically included in outcomes
  - Environmental exposure or condition not included in outcomes
  - Wearable only used to assess the effect of another intervention (eg, cooling)

Individual effects of climate change were limited to those resulting from exposure to weather extremes, as the topic would have been too broad otherwise [4]. As per the 2021-published Intergovernmental Panel on Climate Change report [3], we included exposure to heat and heat waves, heavy precipitation, floods, tropical cyclones, droughts, and wildfires. As heat and heat waves are often defined as extremes relative to the local climate (ie, daily minimum and maximum temperatures above the 95th or 99th percentile of the climatological record or a baseline period) [1,22], we relied on the definitions provided in the included studies. If the authors did not provide a definition, we used one of the following classifications, based on the available data in the screened articles:

- If data were available on wet bulb globe temperature (WBGT) [1,23] or the universal thermal climate index [24], we used >26 °C as a threshold.

- If data were available on ambient temperature and relative humidity, we calculated the heat stress index (HSI) [25] and used a threshold of >26 °C HSI.
- If data were available on ambient temperature, we used the average relative humidity at the study location (city or country) during the study period to calculate the HSI.

Studies on the effect of temperature on sleep were included even for lower ambient temperatures, as previous research has shown that small temperature changes already have adverse effects on sleep quality and duration [26] because humans only have a minimal ability to thermoregulate in rapid eye movement sleep phases [27]. We also included studies that reported on indoor heat exposure in climatic chambers or laboratories. We excluded studies on heat exposure during firefighting and mining, as we considered them job-specific and they predominantly assessed the microclimate inside the protective gear [28].

In case no full text was available or information on the wearables was missing, the authors were contacted 3 times before exclusion.

Search Strategy and Information Sources

The full search was conducted on September 1, 2021, by 1 reviewer (MK) in 6 electronic databases: PubMed (MEDLINE), Scopus, CINAHL (EBSCOhost), IEEE Xplore, Ovid (MEDLINE[R]), and Web of Science. Gray literature was searched with Google Scholar, and the first 1000 search results were included [29]. We manually searched references of relevant included and excluded articles for further sources of evidence.

We followed the Population/Patients, Intervention, Comparison, and Outcome (PICO) framework to compile the search strategy. Population (P) included study participants wearing a wearable. Intervention (I) included exposure to climate change–induced weather extremes. No comparison (C) was required. Outcomes (O) included psychological and physiological health parameters or exposure measurable with wearables. Accordingly, the databases were searched using a search string including synonyms and medical subject headings terms for these concepts. Search strings were adapted to the specific requirements of each database (see Multimedia Appendix 2 for the full search strings). We applied a search filter for publications after January 1, 2010.

Study Selection

The search results were imported into the literature reference management system EndNote 20 (Clarivate Analytics) and then imported into the systematic review management software Covidence (Veritas Health Innovation) where duplicates were removed automatically as well as manually. We screened titles and abstracts, as well as full texts, with application of the inclusion and exclusion criteria (see Textbox 1 for a full list of criteria). Subsequently, we extracted data from the included literature. The screening process was piloted prior with a sample of 20 articles. The literature was screened by 2 independent reviewers (MK and IM). Any disagreements were resolved by consensus between the 2 reviewers (MK and IM) and an independent researcher (SB).

Data Extraction

A data-charting form was developed using the Covidence software template and piloted on 3 articles; data were charted by the 2 reviewers independently, and any disagreements were mutually resolved. The following data categories were extracted and synthesized [17,18]: title, author, year, country of study, objectives of study, demographics of the study population, sample size, methods, intervention type, outcomes, and key findings related to the scoping review question. In addition, the following items were extracted: wearable models, measured parameters with wearable, study setting, climate change–related environmental conditions including the measurement method, and methods used for data analysis or correlation.

Synthesis of Results

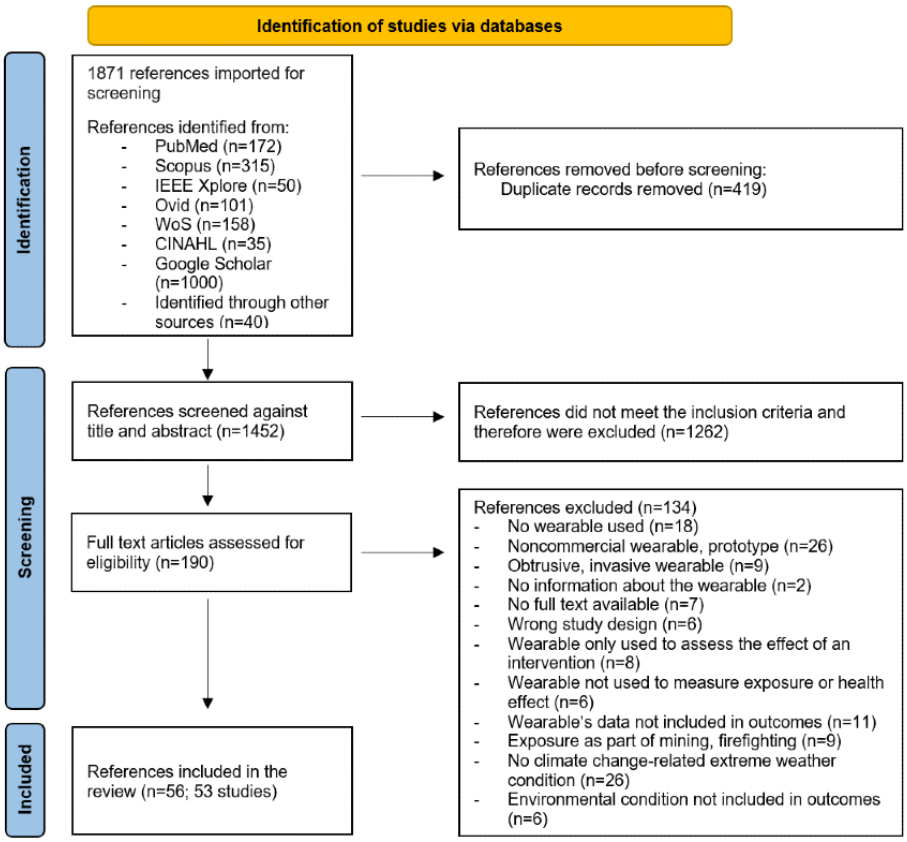
The characteristics of the included studies and the study populations were summarized using Microsoft Excel (version 2206; Microsoft Corporation), and the study outcomes were narratively synthesized. The purpose of the use of the wearables was identified according to three categories: (1) validity and comparison in extreme conditions, (2) measuring individual exposure, or (3) measuring direct health effects.

Results

Overview

The initial search yielded 1831 results, and 40 references were added after a manual search. We removed 419 duplicates and screened the titles and abstracts of the remaining 1452 nonduplicates. From a total of 190 screened full-text articles (186 studies), we included 53 studies (56 articles; see Figure 1 for the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flow diagram) including 1 preprint article [30]. For the conducting of this scoping review the preprint article was used and is therefore cited throughout the manuscript instead of the accepted article [31] that was published after our last search and data extraction process.

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



Study Characteristics

In total, we included a study population of 55,284 participants in this review (the characteristics of the included studies are summarized in Figure 2 and Table 1). Overall, there have been an increasing number of publications using wearables in the context of climate change and health research since 2010 (Table 2). The included studies were mostly observational (35/53, 66%) and crossover studies (21/53, 40%). Most studies were conducted in countries classified by the World Bank in 2022 [32] as upper–middle-income (5/53, 9%) and high-income countries (47/53, 87%), especially with more than half of the total studies conducted in North America (31/53, 58%). A few studies (4/53, 8%) included lower–middle-income countries. Most studies were conducted in urban settings (25/53, 47%) or in a climatic chamber (19/53, 36%), with a short study duration of up to 1 week (16/53, 30%) or up to 5 cross-sectional data collection points (17/53, 32%).

The median number of participants per study was 39 (range 6-47,628), comprising an average of 67% of male participants (Table 3 shows the demographics of the study population). In total, of the 53 studies, 15 (28%) studies focused solely on male participants versus 3 (6%) studies that only included female participants. A few studies (3/53, 6%) specifically included nonhealthy participants. Most study populations consisted of outdoor workers (14/53, 26%), including farm workers, construction workers, traffic police officers, or other workers, as well as the general population (11/53, 21%) or university members (students and staff; 7/53, 13%). Of the 53 studies, 2 (4%) studies included older adults and 4 (8%) studies included children. In addition, the study populations of individuals in the military, athletes, and homeless individuals were each represented in 2% (1/53) of studies.

Figure 2. Map of study locations (countries). Minor et al [31] mentioned 68 countries across all continents (except Antarctica) but did not further specify, so they were not included in the map.

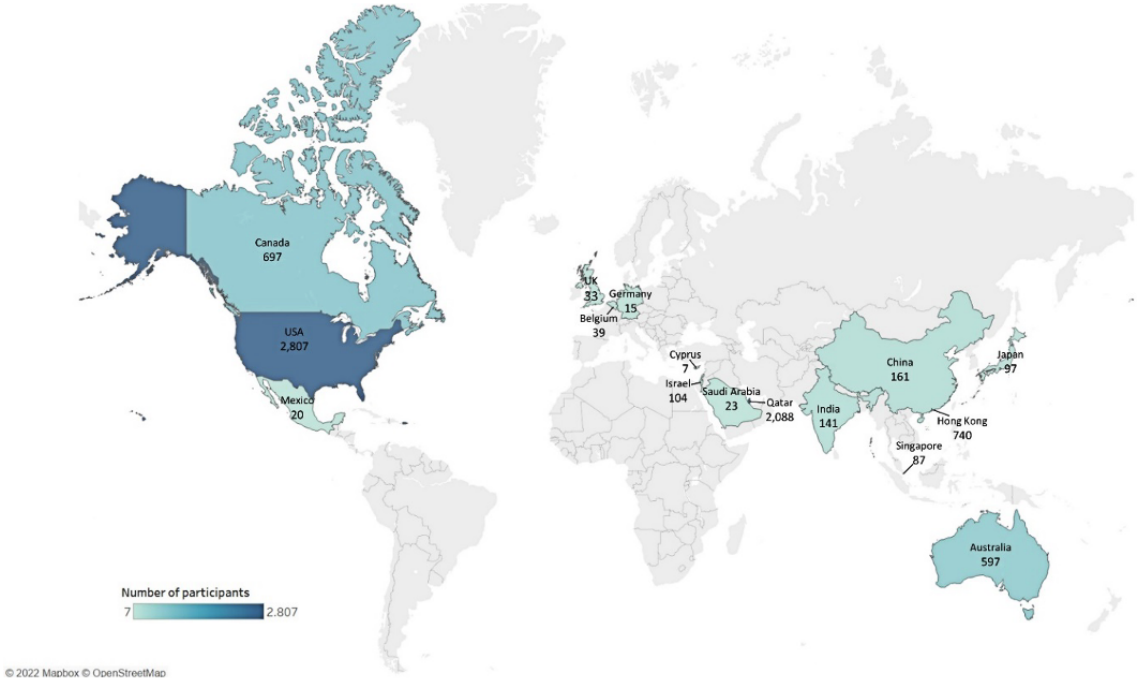


Table 1. Study characteristics.

Study characteristics	Studies (N=53), n (%)	Participants (N=55,284), n (%)
Regions and countries[a]		
North America	30 (56.6)	3524 (6.4)
United States	24 (45.3)	2807 (5.1)
Canada	5 (9.4)	697 (1.3)
Mexico	1 (1.9)	20 (0)
Asia	11 (20.8)	1226 (2.2)
Hong Kong	1 (1.9)	740 (1.3)
China	3 (5.7)	161 (0.3)
India	3 (5.7)	141 (0.3)
Japan	2 (3.8)	97 (0.2)
Singapore	2 (3.8)	87 (0.2)
Europe	5 (9.4)	94 (0.2)
Belgium	1 (1.9)	39 (0.1)
United Kingdom	2 (3.8)	33 (0.1)
Germany	1 (1.9)	15 (0)
Cyprus	1 (1.9)	7 (0)
Oceania	4 (7.5)	597 (1.1)
Australia	4 (7.5)	597 (1.1)
Middle East	3 (5.7)	2192 (4)
Qatar	1 (1.9)	2088 (3. 8)
Israel	1 (1.9)	104 (0.2)
Saudi Arabia	1 (1.9)	23 (0)
South America	0 (0.0)	0 (0)
Africa	0 (0.0)	0 (0)
Countries not specified (68 countries: 42 high-income countries; 17 upper–middle-income countries; 9 lower–middle-income countries)	1 (1.9)	47,628 (86.2)
Study setting[a,b]		
Urban	31 (58.5)	__[c]
Outdoor	9 (17.0)	—
Indoor	5 (9.4)	—
Indoor and outdoor	17 (32.1)	—
Rural	11 (20.8)	—
Outdoor	7 (13.2)	—
Indoor	1 (1.9)	—
Indoor and outdoor	3 (5.7)	—
Climatic chamber or laboratory	19 (35.8)	—
Study duration		
Cross-sectional data collection points (up to 4 hours each)		
≤5 data collection points	17 (32.1)	496 (0.9)
≥6 and ≤10 data collection points	3 (5.7)	174 (0.3)
≥11 and ≤50 data collection points	3 (5.7)	116 (0.2)
Continuous monitoring (at least 1 [work] day)		

Study characteristics	Studies (N=53), n (%)	Participants (N=55,284), n (%)
≤7 days	16 (30.2)	3614 (6.5)
≤1 months	6 (11.3)	171 (0.3)
≤6 months	5 (9.4)	542 (1)
≤2 years	3 (5.7)	50,171 (90.8)
Study design		
Experimental crossover study[d]	7 (13.2)	210 (0.4)
Prospective cohort study	20 (37.7)	49,690 (89.9)
Retrospective cohort study	1 (1.9)	104 (0.2)
Prospective observational crossover study[d]	14 (26.4)	5017 (9.1)
Method comparison or evaluation study	11 (20.8)	263 (0.5)

[a]Multiple characteristics may apply per study.

[b]Information for study settings is not available for all study participants and therefore not summarized here as the number of participants per study setting.

[c]Not available.

[d]Each participant serves as their own control or comparison.

Table 2. Years of publication.

Year of publication	Included publications (N=56), n (%)
2013	3 (5)
2014	6 (11)
2015	6 (11)
2016	5 (9)
2017	5 (9)
2018	6 (11)
2019	8 (14)
2020	11 (20)
2021 (until September 1)	6 (11)

Table 3. Demographics of included studies.

Study	Participants monitored with wearables, n	Study population	Sex (male), %	Age (years)	Ethnicity, %
Al-Bouwarthan et al [33], 2020	23	Construction worker	100	Mean 42.7 (SD 8.8)	—[a]
Al-Mohannadi et al [34], 2016	2088	General population	67	Range 18-65	—
Al Sayed et al [35], 2017	12	Male	100	Mean 24.8 (SD 3.8)	—
Bailey et al [36], 2019	38	University member	50	Group 1: mean 32.6 (SD 13); group 2: mean 21.5 (SD 3)	92% White
Benita et al [37], 2020; Benita and Tuncer [38], 2019	10	University student; female	0	Mean 22.8 (SD 1.5)	—
Benjamin et al [39], 2020	19	Athlete; female	0	Mean 20.6 (SD 1.4)	—
Bernhard et al [40], 2015	81	Outdoor worker or general population	35	Mean 52 (rural), 50.5 (urban), and 44.5 (outdoor worker)	93% Black or African American
Cedeño Laurent et al [41], 2018	44	University student; healthy	51	Mean 20.2 (SD 1.8)	40% White
Cheong et al [42], 2020	9	Older adult	22	Range 65-87	67% White, 11% Black, 11% Hispanic or Latino, and 11% other
Cuddy et al [43], 2013	56	Male	100	Mean 22 (SD 3)	—
Culp and Tonelli [44], 2019	20	Farm worker; male	100	Range 18-65	100% Hispanic
Edwards et al [45], 2015	372	Children (age 3 years at recruitment); healthy	52	Mean 3.4 (SD 0.3)	22% Black or African American
Hamatani et al [46], 2017	13	General population	92	—	—
Hass and Ellis [47], 2019	45	General population	37	Range 18-≥65	64% White and 11% Black or African American
Hondula et al [48], 2020	84	General population	—	—	—
Ioannou et al [49], 2017	7	Farm worker; healthy	71	Male: mean 39 (SD 10.8); female: mean 39.5 (SD 13.4)	—
Jehn et al [50], 2014	15	Clinically stable NYHA II-IV[b] patients with PAH[c]	60	Mean 66.7 (SD 5.2)	—
Kakamu et al [51], 2021	84	Construction worker	100	Mean 48.4 (SD 14)	—
Ketko et al [52], 2014	104	Military; male	100	Range 18-21	—
Kim et al [53], 2013	12	Male	100	Mean 25.5 (SD 4.1)	—
Kuras et al [54], 2015	23	General population	39	Range 25-79	74% White and 26% Black or African American
Lam et al [55], 2021	145	University student (first-year student)	34	Mean 18.1 (range 17-21)	—
Larose et al [56], 2014	60	Male; healthy	100	Mean 45.4 (range 20-70)	—
Lewis et al [57], 2016	1095	Children aged 9-11 years	43	Mean 10.6 (SD 0.4)	—
Li et al [58], 2020	10	Construction worker; healthy; male	100	Mean 39.4 (SD 3.6)	—

Study	Participants monitored with wearables, n	Study population	Sex (male), %	Age (years)	Ethnicity, %
Lisman et al [59], 2014	46	Military or university community member; healthy or previous exertional heat stroke	74	Mean 29.7 (SD 5.9)	—
Longo et al [60], 2017	20	Homeless individual or university student	75	Range 18-60	—
Lundgren et al [61], 2014	77	Outdoor worker	86	—	—
MacLean et al [62], 2020	12	Male; healthy	100	Mean 24.2 (SD 3.7)	—
Minor et al [30], 2020	47,628	General population	69	Age distribution: 19-25, 6%; 25-65, 91%; ≥65, 3%	—
Mitchell et al [63], 2018	587	Farm worker	66	Mean 38.6	98% Latino
Nazarian et al [64], 2021	77	General population	52	Range 18-48	100% Asian
Notley et al [65], 2021	50	Young (18-30) and healthy or older (50-70) and healthy; older and T2D[d] or HTN[e]	100	Mean 50 (SD 17); mean per group: 22 (young), 58 (older), 60 (T2D), and 61 (HTN)	—
Ojha et al [66], 2020	10	University student	70	—	—
Pancardo et al [67], 2015	20	Outdoor worker; healthy	55	Mean 28.6 (range 22-51)	—
Quante et al [68], 2017	669	Adolescents aged 12-14 years	49	Mean 12.9 (SD 0.6)	68% White, 14% Black, 3% Hispanic, 3% Asian, and 13% Other
Raval et al [69], 2018	16	Traffic police worker	100	Range 19-57	—
Ravanelli et al [70], 2016; Ravanelli et al [71], 2015	8	Male; healthy	100	Mean 24 (SD 3)	—
Relf et al [72], 2018	14	Female; healthy	0	Mean 26 (SD 7)	—
Relf et al [73], 2020	19	General population; healthy	79	Mean 41 (SD 23)	—
Rosenthal et al [74], 2020	455	General population	42	—	—
Runkle et al [75], 2019; Sugg et al [76], 2018	35	Outdoor worker	100	Mean 39.2	74% White, 14% Black or African American, 9% Hispanic, and 2% American Indian or Alaska Native
Sahu et al [77], 2013	48	Farm worker	100	Range 25-34	—
Seo et al [78], 2016	12	Male; healthy	100	Group 1: mean 23 (SD 1); group 2: mean 23 (SD 2); group 3: mean 24 (SD 2)	—
Shakerian et al [79], 2021	18	University student	78	Female: mean 24 (SD 3.2); male: mean 24 (SD 2.8)	—
Shin et al [80], 2015	9	Young; healthy	67	Mean 23.3 (SD 4.1)	—
Suwei et al [81], 2019	51	Outdoor worker	35	Mean 42.9 (range 21-60)	96% African American
Uejio et al [82], 2018	50	Outdoor worker	92	Mean 44 (SD 11.1)	59% Black, 39% White, and 2% Hispanic
Van Hoye et al [83], 2014	39	University student; healthy	54	Mean 21.4 (SD 1.41)	—
Williams et al [84], 2019	51	Older adult	43	Mean 65.4	67% White
Xiong et al [85], 2020	48	General population	46	Mean 36 (SD 12)	—

Study	Participants monitored with wearables, n	Study population	Sex (male), %	Age (years)	Ethnicity, %
Zheng et al [86], 2019	740	Adolescent or secondary school student	52	Mean 14.7 (SD 1.6)	100% Asian
Zhu et al [87], 2016	6	General population	50	Males: mean 27.3 (SD 2.5); female: mean 22.3 (SD 1.2)	—

[a]The respective information was missing in the article.

[b]NYHA II-IV: New York Heart Association Functional Classification for heart failure stage II-IV.

[c]PAH: pulmonary arterial hypertension.

[d]T2D: type 2 diabetes.

[e]HTN: hypertension.

Wearable Devices

Most of the included studies used 1 (39/53, 74%) or 2 (12/53, 23%) wearables; a few studies (2/53, 4%) used ≥3 devices (study methods and objectives detailed in Table 4). The 70 wearables in the included studies were from 23 different companies overall with Polar Electro (16/53, 30%), Maxim Integrated (13/53, 25%), and Fitbit (5/53, 9%) providing the most frequently used wearables. The most commonly reported use for wearables was the measurement of HR (30/53, 57%), physical activity (15/53, 28%), or individually experienced temperature (IET; the air temperature surrounding the individuals; 14/53, 26%). Other parameters included sleep (duration, onset, wake time, etc), energy expenditure, skin temperature, electrodermal activity, local sweat rate, respiratory rate, or geoposition. Some wearables measured multiple parameters. The devices were mostly wristbands (25/70, 36%), chest straps (18/70, 25%), clipped to clothing or accessories (15/70, 21%), or directly taped to the skin (5/70, 7%). All included studies additionally used questionnaires and further health parameters (eg, blood pressure, weight, height, and urine samples).

Table 4. Study methods and objectives.

Methods and objectives	Studies (N=53), n (%)
Number of wearables per study	
1	37 (74)
2	12 (23)
≥3	2 (4)
Wearable company (models)[a]	
Polar Electro (RCX3, H7, RS800XC, FT1, FT7, Team 2 [Pro], RS800, RS400, WearLink, Accurex Plus, A300, and M400)	16 (30)
Maxim Integrated (iButton Hygrochron and Thermochron)	13 (25)
Fitbit (Ionic, Charge 2, and Flex)	5 (9)
Medtronic (Zephyr BioHarness)	4 (8)
Philips Respironics (Actical and Actiwatch 2), Onset Corp (HOBO Pendant), and Empatica (E4)	3 (6; each)
Crossbridge Scientific (KuduSmart), Actigraph (GT3X and GT3X+), Intel (Basis Peak Watch), BodyMedia (SenseWear Pro 3), Sony (SmartBand Talk SWR30 and SWR12)	2 (4; each)
Omron Healthcare (HJ-720 ITC pedometer), STATSports (Viper Pod), Microsoft (Band), Garmin (Vivoactive HR), Aipermon (APM), Stayhealthy (RT3), GISupply (LW-360HR), Lifensense (Mambo 2), LASCAR (EL-USB-2-LCD+), Easylog (Easylog), PAL Technologies (activPAL and activPAL3C)	1 (2; each)
Measured parameter with wearable[a]	
Heart rate	30 (57)
Physical activity	15 (28)
Energy expenditure	8 (15)
Skin temperature	12 (23)
Electrodermal activity	5 (9)
Sleep (onset, offset duration, and efficiency)	7 (13)
Individually experienced temperature	14 (26)
Others (local sweat rate, respiratory rate, and GPS location)	7 (13)
Wear location of wearable[a]	
Wristband	25 (47)
Chest strap	18 (34)
Attached to clothing or accessories	15 (28)
Taped to the skin	5 (9)
Other: shirt, back strap, around upper arm, or not specified	8 (15)
Climate change–related extreme weather	
Heat	52 (98)
Wildfire	1 (2)
Measured environmental condition[a]	
Temperature	50 (94)
Relative humidity	40 (75)
Precipitation	7 (13)
Other (wind speed, wet bulb temperature, dry bulb temperature, dew point, mean radiant temperature, barometric pressure, visibility, CO_2 concentration, and air quality)	22 (42)
Measurement location or data source for environmental condition[a]	
Nearest weather station	20 (38)
Sensors placed on study site	18 (34)

Methods and objectives	Studies (N=53), n (%)
Climatic chamber or laboratory	18 (34)
Locally installed weather station	4 (8)
Smartphone sensor	2 (4)
Satellite data	2 (4)
Heat stress measure[a]	
Wet bulb globe temperature	14 (26)
Heat stress index	5 (9)
Humidex	2 (4)
Others (universal thermal climate index, heating or cooling degrees, heat stroke index, heat stress days, heat stress level estimation, heat balance equation, extreme heat degree minutes, and physiological equivalent temperature)	1 (2; each)
None	27 (51)
Method of analysis (statistical test)[a]	
Regression (linear, logistic, and Cox)	16 (30)
Linear mixed effect model	16 (30)
Time-series analysis	1 (2)
t test (2-tailed or 1-tailed)	21 (30)
Correlation (Pearson, Spearman, etc)	13 (25)
ANOVA (one-way, repeated measures, and mixed design)	14 (26)
MANOVA	1 (2)
Nonparametric test (Wilcoxon *U* test and Kruskal-Wallis test)	7 (13)
Chi-square and Fisher Exact Test	4 (8)
Bland Altman plot	5 (9)
Spatial correlation	1 (2)
Cohen kappa	1 (2)
Descriptive analysis only	5 (9)
Study objectives and use of wearables[a]	
Studies measuring the correlation of wearables' data and environmental conditions	
Effect of heat on sleep	7 (13)
Effect of heat on physical activity	7 (13)
Effect of heat on heart rate	10 (19)
Other physical responses to heat	6 (11)
Occupational heat stress	8 (15)
Effect of wildfires on physical activity	1 (2)
Studies measuring the individual experienced temperature and comparing it to local or area measurements	10 (19)
Studies assessing the validity and applicability of wearables for their use in extreme weather	14 (26)

[a]Multiple characteristics may apply per study.

Weather or Climate Data

The primary focus was on the use of wearables to measure physiological responses to heat exposure (52/53, 98%). Of the 53 studies, 1 (2%) study assessed the impact of forest fires on individual activity, and 5 (9%) measured the effect of precipitation on activity in addition to heat. The weather or climate conditions were predominantly assessed using data from the nearest weather station (20/53, 38%), sensors placed on the study site (18/53, 34%), or measured in a climatic chamber or laboratory (18/53, 34%). Others accessed weather data from locally installed weather stations, built-in sensors of participants' smartphones, or satellite data. Besides the primarily focused measurements of temperature, precipitation, and relative humidity, 49% (26/53) of the included studies calculated

different heat stress indices (eg, WBGT, HSI, or universal thermal climate index).

Statistical Analysis

The methods of statistical analysis of wearables' data and correlation to climate or weather data were primarily regression, linear mixed effect models, correlation, ANOVA, and 1- or 2-tailed *t* tests (Table 5). Linear regression models or linear mixed effect models, for example, were often used to correlate IETs and area-level temperature data [35,40,42,48,75], but *t* tests were also used for the comparison between both methods [47,54]. Data sources differed between group-level data and participant-level data [42,54]. The associations of heat exposure and wearables-measured parameters were mostly examined with linear mixed effect models or different regression models (linear, logistic, or Cox), adjusted for age, sex, and education [33,34,39,41,42,45,63,65,68,84-86]. For the comparison of the effect of heat between groups with different characteristics such as sex or age and for the comparison of heat-stress and non–heat-stress days, *t* tests, Chi-square tests, and ANOVAs were used [35,37,43,50,55-57,59,61,63,65,84]. Studies that compared wearables measurements with standard devices applied; in addition to *t* tests and ANOVAs, different correlation coefficients and Bland Altman plots for the appraisal of disagreement [36,72,73,78,83]. Other analysis methods were also used. One study [37,38] spatially correlated different urban environmental exposures and body responses during a 10-minute walk in Singapore. Pattern recognition and parametric tests were used to identify stress hot spots on this walk, and 4 machine learning models were trained to test the predictive power of the immediate environment. Another study used machine learning models for heat strain assessment [79].

Table 5. Study findings regarding the associations of demographic characteristics and heat exposure or physical response.

Finding	Adverse effects on sleep	HR[a] increase	Decrease in physical activity	Skin temperature increase	Occupational heat stress	Higher IET[b]
Age (years)						
Positive association	[30]	[56]	[34]	—[c]	[63]	[40]
Null association	—	[65]	—	—	—	—
Sex (female)						
Positive association	[30]	[59]	[34]	—	[61,63]	—
BMI or body fat percentage						
Positive association	—	[59]	—	[44]	[63,75,76]	[40]
Education						
Positive association	—	—	—	—	[75,76]	—
Negative association	—	—	—	—	—	[47]
Income						
Negative association	—	—	—	—	—	[40,47,88]
Homelessness						
Positive association	—	—	—	—	—	[60]
Health status (hypertension and type 2 diabetes)						
Null association	—	[65]	—	—	—	—
Lower-income country						
Positive association	[30]	—	—	—	—	—
From the Eastern Mediterranean region						
Positive association	—	—	[34]	—	—	—

[a]HR: heart rate.

[b]IET: individually experienced temperature.

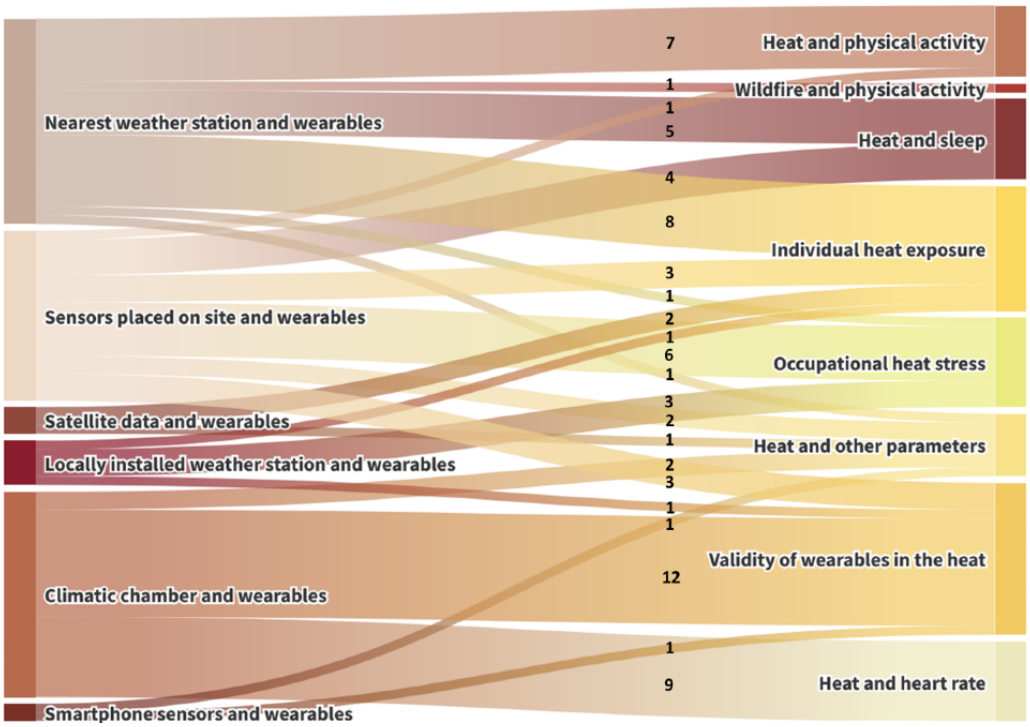
[c]No findings regarding an association were stated in the included studies.

Study Outcomes and Findings Regarding the Use of Wearables in Extreme Weather Conditions

Overview

We categorized the studies according to the use of wearables in extreme weather conditions (Table 4). An overview of the data collection methods for each category is shown in Figure 3. Table 5 displays the reported associations of participants' demographic characteristics and individual exposure or physiological responses to heat. The study findings are summarized in the subsequent sections (see Multimedia Appendix 3 for a detailed compilation).

Figure 3. The Sankey diagram shows the data collection methods that were used for each study outcome. The methods are displayed on the left (weather or climate measurement method and wearable) and connected to the respective study outcome shown on the right. The numbers show the number of studies that are represented by each link. One study might have more than one study outcome and therefore could be represented in multiple strings.



Correlation of Wearables' Data and Extreme Weather

Effect of Heat on Sleep

Of the 53 studies, 7 (13%) studies [30,41,68,80,84,85,87] examined the effect of bedroom environmental conditions on sleep parameters (efficiency, rapid eye movement sleep, duration, sleep onset latency, and sleep disruptions) and all 7 studies found a negative correlation between higher ambient temperature and sleep in their study cohorts. One large-scale study [30] found significantly larger negative effects of heat on sleep duration for residents from lower-income countries, older adults, and females, with no evidence of short-term acclimatization. In contrast, another study [41] found evidence of short-term heat acclimatization in a cohort of young adults.

Effect of Heat on Physical Activity

A total of 13% (7/53) of studies [34,39,45,50,57,68,86] examined the effect of heat on physical activity (mostly measured in the form of steps), and most (6/7, 86%) studies [34,39,45,50,57,68] found a negative correlation in the general population, children, female soccer players, and patients with pulmonary arterial hypertension and clinically stable heart insufficiency (New York Heart Association Classification Stage II-IV). In contrast, one study [86] found a significant positive correlation for temperatures between 13 °C and 31 °C in a cohort of children on weekend days. One study [34] found that the decrease in physical activity was greater with age in female participants and participants from the Eastern Mediterranean region; Five studies [34,45,57,68,86] examined the effects of precipitation and heat on physical activity, and all (5/5, 100%) found a negative correlation between precipitation and physical activity. Four of these studies [45,57,68,86] observed this effect in cohorts of children.

Effect of Heat on HR

Of the 53 studies, 10 (19%) studies [35,43,52,56,58,59,65,66,70,71,83] examined the effect of heat on HR, and most (7/10, 70%) observed increasing HR in hotter and more humid conditions, especially in older adults. Four studies [35,56,66,83] found no significant effect of hot and humid conditions on HR during exercise among young adults, whereas 1 study [58] found no significant effect in middle-aged participants. In contrast, 1 study [65] found no significant difference in HR response to exercise under heat exposure for different age groups or participants with hypertension or type 2 diabetes. Two studies [43,52] conducting heat exposure tests measured significantly higher, steadily increasing HR for participants classified as "at risk" or "heat intolerant." One study [59] found that BMI, percent body fat, sex, and maximal oxygen uptake were associated with elevated HR during heat tolerance testing. To mitigate these heat effects, 1 study [70,71] found that placing an electric fan 1 m in front of the participants could significantly delay HR increase in hot and humid conditions.

Other Physiological (and Psychological) Responses to Heat

Of the 53 studies, 6 (11%) studies [37,38,42,55,66,73,83] examined further body responses to heat exposure, including increasing skin temperature, electrodermal activity, skin conductance response, and energy expenditure (during high-intensity exercise). One study [37,38] discovered stress hotspots during a route through the city, which may be explained by changes in the immediate environment, such as the transition from a park to a residential area or an exposed area without shade. Another study [42] found a correlation among higher HR, near-body temperature, and outside temperature when participants reported mild anxiety.

Two studies observed signs of short-term heat acclimatization over 9 to 10 days in the form of increasing local sweat rate [73] and decreasing HRs and metabolic rates for both local and nonlocal students. The differences between both groups (higher values for nonlocal students) were assimilated in the second week [55].

Occupational Heat Stress

Of the 53 studies, 8 (15%) studies [33,44,49,51,61,63,75-77] used wearables to investigate the physiological effects of occupational heat stress on outdoor workers, including construction, farm, and ground management workers. Six studies [44,49,61,63,75-77] found an association between occupational heat exposure and physiological responses, including increasing HR [44,75-77], metabolic rate [61], skin temperature [44,49], and decreasing physical activity [63]. One study [33] found that WBGT and heat stress exposure were stronger predictors of cardiovascular strain (measured as HR reserve) than energy expenditure during construction work. Five studies found associations between demographic characteristics and physical responses to occupational heat stress, including the female sex (2/8, 25%) [61,63], older age (1/8, 13%) [63], higher BMI (2/8, 25%) [44,75,76], and education (1/8, 13%) [75,76]. Furthermore, of the 8 studies, in 1 study [75,76], the perception of heat as an occupational hazard and officially issued heat alerts were associated with a lower heat strain risk. Two studies [33,61] observed that occupational heat stress exposure frequently reached critical conditions.

Effect of Air Quality During Wildfires on Physical Activity

Of the 53 studies, 1 (2%) study [74] found a statistically significant reduction in daily step counts with progressively worse air quality during wildfires, with an 18% reduction in daily step count when the air quality index exceeded 200 (considered very unhealthy, and public health warnings were typically issued) compared with less than 100 (considered good air quality).

Individual Heat Exposure and Comparison to Area-Level Measurements

Of the 53 studies, 10 (19%) studies compared individual heat exposure measured with wearables to area-level measurements [36,40,42,47,48,54,60,69,81,82]. Individual heat exposure was measured in the form of IET (the air temperature surrounding the individuals).

Three studies [42,48,54] found high heterogeneity in IETs, indicating interindividual differences in time spent outdoors, 2 of which [42,48] found little to no correlation between temperature measured at the nearest weather station and IET while participants spent their time indoors and outdoors. Two studies [36,54] found different associations between IETs and weather station measurements between daytime and nighttime. Three studies [69,81,82] observed higher IETs when compared with measurements from the nearest weather station or the locally installed weather station for outdoor workers [69,81,82]; individual exposure frequently exceeded the recommended values. In contrast, 4 studies [40,47,48,54] found that average temperature measurements from the nearest weather station were higher than average daily IETs, including heat wave periods compared with non–heat wave periods (2/10, 20%) [47,54]. Four studies [40,47,48,60] found associations between demographic characteristics and IET, including a negative association of income (3/10, 30%) [40,47,48], education (1/10, 10%) [47], homelessness (1/10, 10%) [60], urban environment (1/10, 10%) [40], and higher body fat percentage (1/10, 10%) [40] with heat exposure (especially indoors and during the night), whereas older study participants experienced higher strain due to nighttime exposure compared with the comparison groups (1/10, 10%) [40].

Validity and Reliability of Wearables in Extreme Weather

Of the 53 studies, a total of 14 (26%) studies [35,36,43,46,52,53,62,64,67,72,73,78-80,83] evaluated and compared different wearables and their validity and reliability in the context of heat exposures.

Seven different wearables were compared with gold-standard methods for measuring various physiological parameters during heat exposure and under different levels of physical activity. Five studies found no significant differences in measurements between wearables and gold-standard devices, including the Polar chest strap [35], Hexoskin shirt [35], and Zephyr BioHarness [53] for HR measurements, the SenseWear Pro 3 and Actiwatch 2 wearables [80] for the measurement of sleep, 2 different SenseWear Algorithms for the estimation of energy expenditure during low-intensity activity [83], and the KuduSmart [72,73] for local sweat rate measurements. Two studies [36,62] compared wearables and their placements using standard measurements. One study [62] concluded that single-location skin temperature measurements at the chest, scapula, and thigh with iButton wearables taped to the skin were the only positions to agree with mean skin temperature (standard method) under all conditions and that wearables outperformed the infrared device. The other study [36] found no significant difference between sensor types (iButton Thermochron and Hygrochron, HOBO Pendant) and placements (on shirt collar, shoe, or backpack) for the measurement of IET but the correlation between sensors was the lowest during high-intensity activities.

Three studies (3/14, 21%) [46,64,78] compared different methods of using wearables data for core temperature estimation to measured values and found good overall results using HR [78]; HR and skin temperature [46]; and HR, skin temperature, and near-body temperature [64]. Four studies [43,52,67,79] evaluated different methods to use wearables data (HR or cardiac cost, electrodermal activity, and skin temperature) to assess heat strain and all found high sensitivity or accuracy. Two of these studies paired wearables data with skin temperature or rectal and core temperature measurements for assessment.

Discussion

Principal Findings

Overall, the included studies revealed a diverse spectrum of wearable devices, with a particular emphasis on physical responses to heat. Heat was found to adversely affect sleep, physical activity, occupational stress exposure, and other physical and psychological parameters. Air quality during

wildfires was another weather extreme examined in 1 study and was found to negatively affect physical activity. A comparison of individual exposures against weather station area-level measurements showed high differences. Wearables were found to provide valid and reliable metrics for assessing physiological responses in extreme weather conditions. We identified a slight increase in the number of scientific publications in recent years.

Study Settings

The vast majority of studies were conducted in upper–middle-income and high-income countries, as has been reported in prior publications [2]. Half of the study participants in the included studies were from North America, even though sub-Saharan Africa, South and Central America, and parts of Asia are projected to be the regions most affected by climate change [12]. Different obstacles to the use of wearables in low-income countries could be the reason for this finding. From the participants' perspective, acceptability of and adherence to the use of wearables could be affected by skepticism [89] or fear of theft and loss of the device [90], since these populations are often not as exposed to these technologies as study populations from high-income countries [91,92]. From the researchers' perspective, another obstacle could be the lack of smartphones that are needed for the connection of most wearables and lack of high-speed internet connections [90,91] and therefore higher study costs. Some studies [30,74] have included participants who already owned wearables, which translates into higher acceptability by potential study participants. However, this approach is hardly possible in populations where wearables are not commonly used [92]. A few studies are ongoing in low-income countries, among others is a study exploring the feasibility of consumer-grade wearable devices in Burkina Faso and Kenya [93]. However, not only wearables data present an obstacle for studies in low-resource contexts but also weather and climate data are not as widely available with less granular spatial distribution of weather stations, especially in remote regions [94]. Using other systems to conduct weather data, such as small, portable sensor systems, has been found to be a possibility for low-income countries [95], and within the framework of some studies, sensors or weather stations were installed in the study region [36,49,63,75,76]. Many of the included studies were conducted in laboratory settings or ran over a short study duration with few participants. Laboratory settings do not necessarily reflect real life and cause low ecological validity. We found only a few studies [30,34,57] that conducted large-scale studies in real-life settings of the participants. By contacting wearable users and collecting their data [30,96], large-scale studies over long periods with population sizes as big as half a million can be conducted quite easily. When the conducted data are correlated with available weather data, important insights into the continuous and long-term health effects of climate change can be gained.

Study Populations

The study populations of the included studies primarily comprised healthy participants and only few that included cohorts with vulnerable populations, such as patients with diabetes or heart insufficiency [50,65]. Wearables have previously been used to assess health in cohorts of patients with chronic diseases, such as cardiovascular diseases [97], and could provide an opportunity to better assess vulnerability to climate change in patients with pre-existing conditions. Different age groups and the association between age and vulnerability to heat have been examined in some studies, mainly showing a higher negative impact of heat on sleep, physical activity, and HR for older participants, as well as higher exposure to nighttime heat. In addition, other demographic characteristics, such as higher BMI and lower income, were associated with heat exposure and larger health effects in some studies. These findings on immediate and short-term health effects are consistent with those of other studies that assessed severe health outcomes, such as hospitalization rates [5]. However, not all studies found similar results; for example, Notley et al [65] found no significant differences between age groups or between healthy and nonhealthy participants. Overall, more research is needed to understand the causal relationship between different population characteristics and the health effects of climate change or climate change–induced weather extremes.

Wearables

The range of included wearables and measurements was broad; however, most studies only used HR, temperature, and accelerometry data. In addition, the validity of various wearables was confirmed to be high in extreme conditions, such as heat, including heat strain assessment, and to be an accepted, noninvasive method for evaluating core temperature.

We also found that many studies used iButtons from Maxim Integrated for ambient and skin temperature measurements; some studies taped the devices to the skin with medical tape. However, we could not find any information from the manufacturer on whether their devices are suitable for measuring skin temperature in this manner.

Weather or Climate Data

Our findings indicate that most studies relied on weather station data, with weather stations frequently located outside cities (eg, airports), and that most studies provided no information on the distance between the weather stations and the study population. However, most studies comparing IETs and area-level measurements from the nearest weather station have shown that they often do not accurately represent individual heat exposure. Future studies should consider individual-level measurements for a more exact heat prediction that captures the effects of time spent indoors and heat adaptation measures such as air conditioning. Almost all of the research has addressed one climate change–related weather extreme: heat. Although heat poses one of the most immediate health threats of climate change [1], other weather extremes and their health effects should not be neglected.

Health Effects of Extreme Weather

Most of the included studies showed an association of weather extremes with adverse health outcomes, particularly for heat. The adverse health effects of heat included less sleep (quality), less physical activity, increased HR, and higher skin temperature. A few studies have also observed changes in physical response to heat, indicating short-term heat adaptation

[41,55]. The heat effects are consistent with those of prior studies that did not use wearables [5,26,98-100]. Studies examining occupational heat stress using wearables found physical effects similar to those mentioned above for other study populations, including cardiac strain and decreasing physical activity [44,49,61,63,75-77]. The physical effects of occupational heat stress are extensive and may impose a significant economic burden owing to the decreased working capacity [101]. The IET by outdoor workers is often higher than area-level measurements and frequently reaches critical thermal conditions during work [33,61], which is important to consider for the assessment of occupational safety.

Individual Heat Exposure

The included studies assessed individual heat exposure with IET, which mostly did not align with area-level measurements and showed high interindividual variances. Sociodemographic factors such as age, income, and education were found to be associated with higher heat exposure, potentially due to a lack of access to cooling methods, such as air conditioning, which have been shown to successfully mitigate the adverse effects of heat [42]. Lower IETs were found during heat waves when heat warnings were made publicly available compared with non–heat wave periods, emphasizing the importance of publicly available information as a means of mitigating individual heat adaptation strategies. In some studies [54,84], indoor temperatures during the night were higher than outdoor temperatures, even with access to air conditioning. This could be explained by the slower cooling rates indoors owing to heat storage in buildings [102]. However, it is important to note that body heat might affect IET measurements by wearables worn on the body or clothes. This aspect was not considered in the included articles.

Limitations

One of the limitations of this review is the exclusion of prototype wearables. Even though studies with prototypes are hardly reproducible, they (especially large-scale studies) may provide valuable insights into the direct health effects of climate change that have not been considered in this review. In addition, we had to exclude many studies because they used invasive or obtrusive wearables. These devices may affect the participants' compliance and make conducting a study outside the laboratory challenging. Second, with the limitation of climate change–related conditions to extreme weather events, we excluded other effects of climate change as well as moderate effects. Other studies and reviews have examined the effects of seasons and moderate weather conditions on health [103,104]. Furthermore, the definitions of extreme weather conditions in our review, especially heat, were not consistent; therefore, comparisons between the study outcomes must be considered cautiously. Third, we excluded studies that did not use wearables to measure the effects of extreme weather, but the effects of interventions, such as studies comparing the effects of fan use and air conditioning. Studies that assessed interventions against heat stress [105,106] could provide important insights into how to best prevent the adverse health effects of climate change.

Conclusions

We found a broad range of wearables to be used in the context of climate change and health research. The validity of many wearables compared with standard devices or methods is high, even in extreme heat. The included studies found that the effects of extreme weather conditions on health can be examined and correlated with wearable data. They showed that heat has adverse effects on wearables-measured variables such as HR, physical activity, and sleep. The findings have underlined individual factors to be associated with higher vulnerability. Furthermore, wearables have been demonstrated to be a suitable tool for assessing individual heat exposure. This could be especially valuable if other meteorological data are not available or during exposure times with large disparities between individual- and area-level measures, such as heat waves or at night. We have identified gaps in the research regarding the use of wearables in low-income contexts and for long-term observations in large-scale studies. For further research, wearables may be a valuable method to generate insights and data at the individual level to better understand the impact of climate change on health, including moderate and short-term effects. As a next step, wearables could be used for the evaluation of adaptation measures.

Acknowledgments

The authors wish to thank the German Research Foundation (Deutsche Forschungsgemeinschaft) for supporting this work as part of a Deutsche Forschungsgemeinschaft–funded research unit (Forschungsgruppe FOR 2936/project: 409670289). We acknowledge financial support by the Else Kröner-Fresenius-Stiftung within the Heidelberg Graduate School of Global Health, by the German Research Foundation (Deutsche Forschungsgemeinschaft) within the funding programme Open Access Publishing, by the Baden-Württemberg Ministry of Science, Research and the Arts and by Ruprecht-Karls-Universität Heidelberg. The funders did not have a role in the design, data collection and analysis, decision to publish, or writing of the manuscript.

Authors' Contributions

MK and SB were responsible for the conceptualization and design of this work. MK developed the search queries and performed the search. Literature screening and data extraction were completed by MK and IM. MK drafted the original manuscript with significant contribution from all authors in editing and revisions. The final manuscript was approved by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 554 KB - mhealth_v10i9e39532_app1.pdf]

Multimedia Appendix 2
Search strings for each database.
[DOCX File , 14 KB - mhealth_v10i9e39532_app2.docx]

Multimedia Appendix 3
Information about the wearables and weather or climate measures for each study.
[DOCX File , 56 KB - mhealth_v10i9e39532_app3.docx]

References

1. Watts N, Amann M, Arnell N, Ayeb-Karlsson S, Beagley J, Belesova K, et al. The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises. The Lancet 2021 Jan 09;397(10269):129-170. [doi: 10.1016/S0140-6736(20)32290-X] [Medline: 33278353]
2. Berrang-Ford L, Sietsma AJ, Callaghan M, Minx JC, Scheelbeek PF, Haddaway NR, et al. Systematic mapping of global research on climate and health: a machine learning review. Lancet Planet Health 2021 Aug;5(8):e514-e525 [FREE Full text] [doi: 10.1016/S2542-5196(21)00179-0] [Medline: 34270917]
3. Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, et al. IPCC 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. 2021. URL: https://www.ipcc.ch/report/ar6/wg1/ [accessed 2022-02-14]
4. Smith KR, Woodward A, Campbell-Lendrum D, Chadee DD, Honda Y, Liu Q, et al. Human health: impacts, adaptation, and co-benefits. In: Field CB, Barros VR, Dokken DJ, Mach KJ, Mastrandrea MD, Bilir TE, et al, editors. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2014:709-754.
5. Rocque RJ, Beaudoin C, Ndjaboue R, Cameron L, Poirier-Bergeron L, Poulin-Rheault RA, et al. Health effects of climate change: an overview of systematic reviews. BMJ Open 2021 Jun 09;11(6):e046333 [FREE Full text] [doi: 10.1136/bmjopen-2020-046333] [Medline: 34108165]
6. Harlan SL, Brazel AJ, Prashad L, Stefanov WL, Larsen L. Neighborhood microclimates and vulnerability to heat stress. Soc Sci Med 2006 Dec;63(11):2847-2863. [doi: 10.1016/j.socscimed.2006.07.030] [Medline: 16996668]
7. White-Newsome JL, Sánchez BN, Jolliet O, Zhang Z, Parker EA, Dvonch JT, et al. Climate change and health: indoor heat exposure in vulnerable populations. Environ Res 2012 Jan;112:20-27 [FREE Full text] [doi: 10.1016/j.envres.2011.10.008] [Medline: 22071034]
8. Zapata-Lamana R, Lalanza JF, Losilla JM, Parrado E, Capdevila L. mHealth technology for ecological momentary assessment in physical activity research: a systematic review. PeerJ 2020 Mar 26;8:e8848 [FREE Full text] [doi: 10.7717/peerj.8848] [Medline: 32257648]
9. Kristoffersson A, Lindén M. A systematic review on the use of wearable body sensors for health monitoring: a qualitative synthesis. Sensors (Basel) 2020 Mar 09;20(5):1502 [FREE Full text] [doi: 10.3390/s20051502] [Medline: 32182907]
10. Nazarian N, Lee JK. Personal assessment of urban heat exposure: a systematic review. Environ Res Lett 2021 Feb 26;16(3):033005. [doi: 10.1088/1748-9326/abd350]
11. Notley SR, Flouris AD, Kenny GP. On the use of wearable physiological monitors to assess heat strain during occupational heat stress. Appl Physiol Nutr Metab 2018 Sep;43(9):869-881. [doi: 10.1139/apnm-2018-0173] [Medline: 29726698]
12. Pörtner HO, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, et al. IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. 2022. URL: https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_FullReport.pdf [accessed 2022-03-08]
13. 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]
14. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
15. Austad H, Wiggen Ø, Færevik H, Seeberg TM. Towards a wearable sensor system for continuous occupational cold stress assessment. Ind Health 2018 Jun 01;56(3):228-240 [FREE Full text] [doi: 10.2486/indhealth.2017-0162] [Medline: 29353859]

16. Pham S, Yeap D, Escalera G, Basu R, Wu X, Kenyon NJ, et al. Wearable sensor system to monitor physical activity and the physiological effects of heat exposure. Sensors (Basel) 2020 Feb 06;20(3):855 [FREE Full text] [doi: 10.3390/s20030855] [Medline: 32041097]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
20. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
21. Salamone F, Masullo M, Sibilio S. Wearable devices for environmental monitoring in the built environment: a systematic review. Sensors (Basel) 2021 Jul 10;21(14):4727 [FREE Full text] [doi: 10.3390/s21144727] [Medline: 34300467]
22. de Perez EC, van Aalst M, Bischiniotis K, Mason S, Nissan H, Pappenberger F, et al. Global predictability of temperature extremes. Environ Res Lett 2018 May 01;13(5):054017. [doi: 10.1088/1748-9326/aab94a]
23. Parsons K. Heat stress standard ISO 7243 and its global application. Ind Health 2006 Jul;44(3):368-379 [FREE Full text] [doi: 10.2486/indhealth.44.368] [Medline: 16922180]
24. Di Napoli C, Pappenberger F, Cloke HL. Verification of heat stress thresholds for a health-based heat-wave definition. J Appl Meteorol Climatol 2019 Jun 1;58(6):1177-1194. [doi: 10.1175/jamc-d-18-0246.1]
25. What is the heat index? US Department of Commerce National Weather Service. URL: https://www.weather.gov/ama/heatindex [accessed 2021-12-22]
26. Rifkin DI, Long MW, Perry MJ. Climate change and sleep: a systematic review of the literature and conceptual framework. Sleep Med Rev 2018 Dec;42:3-9 [FREE Full text] [doi: 10.1016/j.smrv.2018.07.007] [Medline: 30177247]
27. Hirshkowitz M. Normal human sleep: an overview. Med Clin North Am 2004 May;88(3):551-vii. [doi: 10.1016/j.mcna.2004.01.001] [Medline: 15087204]
28. McQuerry M. Wildland firefighting and wearable technology: a review. J Textile Eng Fash Technol 2020 Feb 19;2(1):1-5.
29. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015 Sep 17;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
30. Minor K, Bjerre-Nielsen A, Jonasdottir S, Lehmann S, Obradovich N. Ambient heat and human sleep. arXiv 2020 Nov 13.
31. Minor K, Bjerre-Nielsen A, Jonasdottir SS, Lehmann S, Obradovich N. Rising temperatures erode human sleep globally. One Earth 2022 May;5(5):534-549. [doi: 10.1016/j.oneear.2022.04.008]
32. World Bank Country and Lending Groups. The World Bank. 2022. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2022-02-21]
33. Al-Bouwarthan M, Quinn MM, Kriebel D, Wegman DH. A field evaluation of construction workers' activity, hydration status, and heat strain in the extreme summer heat of Saudi Arabia. Ann Work Expo Health 2020 Jun 24;64(5):522-535. [doi: 10.1093/annweh/wxaa029] [Medline: 32219304]
34. Al-Mohannadi AS, Farooq A, Burnett A, Van Der Walt M, Al-Kuwari MG. Impact of climatic conditions on physical activity: a 2-year cohort study in the Arabian Gulf region. J Phys Act Health 2016 Sep;13(9):929-937. [doi: 10.1123/jpah.2015-0593] [Medline: 27631650]
35. Al Sayed C, Vinches L, Hallé S. Validation of a wearable biometric system’s ability to monitor heart rate in two different climate conditions under variable physical activities. E-Health Telecommun Syst Network 2017 Jun;6(2):19-30. [doi: 10.4236/etsn.2017.62002]
36. Bailey E, Fuhrmann C, Runkle J, Stevens S, Brown M, Sugg M. Wearable sensors for personal temperature exposure assessments: a comparative study. Environ Res 2020 Jan;180:108858. [doi: 10.1016/j.envres.2019.108858] [Medline: 31708175]
37. Benita F, Bansal G, Virupaksha D, Scandola F, Tunçer B. Body responses towards a morning walk in a tropical city. Landsc Res 2020 Sep 08;45(8):966-983. [doi: 10.1080/01426397.2020.1808956]
38. Benita F, Tunçer B. Exploring the effect of urban features and immediate environment on body responses. Urban For Urban Green 2019 Jul;43:126365. [doi: 10.1016/j.ufug.2019.126365]
39. Benjamin CL, Hosokawa Y, Curtis RM, Schaefer DA, Bergin RT, Abegg MR, et al. Environmental conditions, preseason fitness levels, and game workload: analysis of a female NCAA DI national championship soccer season. J Strength Cond Res 2020 Apr;34(4):988-994. [doi: 10.1519/JSC.0000000000003535] [Medline: 31996612]
40. Bernhard MC, Kent ST, Sloan ME, Evans MB, McClure LA, Gohlke JM. Measuring personal heat exposure in an urban and rural environment. Environ Res 2015 Feb;137:410-418 [FREE Full text] [doi: 10.1016/j.envres.2014.11.002] [Medline: 25617601]

41. Cedeño Laurent JG, Williams A, Oulhote Y, Zanobetti A, Allen JG, Spengler JD. Reduced cognitive function during a heat wave among residents of non-air-conditioned buildings: an observational study of young adults in the summer of 2016. PLoS Med 2018 Jul 10;15(7):e1002605 [FREE Full text] [doi: 10.1371/journal.pmed.1002605] [Medline: 29990359]
42. Cheong SM, Bautista C, Ortiz L. Sensing physiological change and mental stress in older adults from hot weather. IEEE Access 2020 Mar 20;8:70171-70181. [doi: 10.1109/access.2020.2982153]
43. Cuddy JS, Buller M, Hailes WS, Ruby BC. Skin temperature and heart rate can be used to estimate physiological strain during exercise in the heat in a cohort of fit and unfit males. Mil Med 2013 Jul;178(7):e841-e847. [doi: 10.7205/MILMED-D-12-00524] [Medline: 23820362]
44. Culp K, Tonelli S. Heat-related illness in midwestern Hispanic farmworkers: a descriptive analysis of hydration status and reported symptoms. Workplace Health Saf 2019 Apr;67(4):168-178 [FREE Full text] [doi: 10.1177/2165079918813380] [Medline: 30724664]
45. Edwards NM, Myer GD, Kalkwarf HJ, Woo JG, Khoury PR, Hewett TE, et al. Outdoor temperature, precipitation, and wind speed affect physical activity levels in children: a longitudinal cohort study. J Phys Act Health 2015 Aug;12(8):1074-1081 [FREE Full text] [doi: 10.1123/jpah.2014-0125] [Medline: 25423667]
46. Hamatani T, Uchiyama A, Higashino T. HeatWatch: preventing heatstroke using a smart watch. In: Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops. 2017 Presented at: PerComW '17; March 13-17, 2017; Kona, HI, USA p. 661-666. [doi: 10.1109/percomw.2017.7917642]
47. Hass AL, Ellis KN. Using wearable sensors to assess how a heatwave affects individual heat exposure, perceptions, and adaption methods. Int J Biometeorol 2019 Dec;63(12):1585-1595. [doi: 10.1007/s00484-019-01770-6] [Medline: 31414184]
48. Hondula DM, Kuras ER, Betzel S, Drake L, Eneboe J, Kaml M, et al. Novel metrics for relating personal heat exposure to social risk factors and outdoor ambient temperature. Environ Int 2021 Jan;146:106271 [FREE Full text] [doi: 10.1016/j.envint.2020.106271] [Medline: 33395929]
49. Ioannou LG, Tsoutsoubi L, Samoutis G, Bogataj LK, Kenny GP, Nybo L, et al. Time-motion analysis as a novel approach for evaluating the impact of environmental heat exposure on labor loss in agriculture workers. Temperature (Austin) 2017 Jul 12;4(3):330-340 [FREE Full text] [doi: 10.1080/23328940.2017.1338210] [Medline: 28944274]
50. Jehn M, Gebhardt A, Liebers U, Kiran B, Scherer D, Endlicher W, et al. Heat stress is associated with reduced health status in pulmonary arterial hypertension: a prospective study cohort. Lung 2014 Aug;192(4):619-624. [doi: 10.1007/s00408-014-9587-4] [Medline: 24816966]
51. Kakamu T, Endo S, Hidaka T, Masuishi Y, Kasuga H, Fukushima T. Heat-related illness risk and associated personal and environmental factors of construction workers during work in summer. Sci Rep 2021 Jan 13;11(1):1119 [FREE Full text] [doi: 10.1038/s41598-020-79876-w] [Medline: 33441683]
52. Ketko I, Eliyahu U, Epstein Y, Heled Y. The thermal-circulatory ratio (TCR): an index to evaluate the tolerance to heat. Temperature (Austin) 2014 Jul 2;1(2):101-106 [FREE Full text] [doi: 10.4161/temp.29752] [Medline: 27583291]
53. Kim JH, Roberge R, Powell JB, Shafer AB, Jon Williams W. Measurement Accuracy of Heart Rate and Respiratory Rate during Graded Exercise and Sustained Exercise in the Heat Using the Zephyr BioHarnessTM. Int J Sports Med 2013;34(6):497-501 [FREE Full text] [doi: 10.1055/s-0032-1327661]
54. Kuras ER, Hondula DM, Brown-Saracino J. Heterogeneity in individually experienced temperatures (IETs) within an urban neighborhood: insights from a new approach to measuring heat exposure. Int J Biometeorol 2015 Oct;59(10):1363-1372. [doi: 10.1007/s00484-014-0946-x] [Medline: 25567543]
55. Lam CK, Hang J, Zhang D, Wang Q, Ren M, Huang C. Effects of short-term physiological and psychological adaptation on summer thermal comfort of outdoor exercising people in China. Building and Environment 2021 Jul;198:107877. [doi: 10.1016/j.buildenv.2021.107877]
56. Larose J, Boulay P, Wright-Beatty HE, Sigal RJ, Hardcastle S, Kenny GP. Age-related differences in heat loss capacity occur under both dry and humid heat stress conditions. J Appl Physiol (1985) 2014 Jul 01;117(1):69-79 [FREE Full text] [doi: 10.1152/japplphysiol.00123.2014] [Medline: 24812643]
57. Lewis LK, Maher C, Belanger K, Tremblay M, Chaput JP, Olds T. At the mercy of the gods: associations between weather, physical activity, and sedentary time in children. Pediatr Exerc Sci 2016 Feb;28(1):152-163. [doi: 10.1123/pes.2015-0076] [Medline: 26098393]
58. Li Y, Du C, Yao R, Li G, Li B. A method to identify individually physiological response differences to heat exposure using Comprehensive Deviation Coefficient (CDC). Energy Build 2020 Jun;217:110003. [doi: 10.1016/j.enbuild.2020.110003]
59. Lisman P, Kazman JB, O'Connor FG, Heled Y, Deuster PA. Heat tolerance testing: association between heat intolerance and anthropometric and fitness measurements. Mil Med 2014 Nov;179(11):1339-1346. [doi: 10.7205/MILMED-D-14-00169] [Medline: 25373064]
60. Longo J, Kuras E, Smith H, Hondula DM, Johnston E. Technology use, exposure to natural hazards, and being digitally invisible: implications for policy analytics. Policy Internet 2017 Mar;9(1):76-108. [doi: 10.1002/poi3.144]
61. Lundgren K, Kuklane K, Venugopal V. Occupational heat stress and associated productivity loss estimation using the PHS model (ISO 7933): a case study from workplaces in Chennai, India. Glob Health Action 2014 Nov 4;7:25283 [FREE Full text] [doi: 10.3402/gha.v7.25283] [Medline: 25373413]

62. MacLean BL, MacLean K, Stewart IB, Hunt AP. Monitoring heat strain: the effect of sensor type and location on single-site and mean skin temperature during work in the heat. Int Arch Occup Environ Health 2021 Apr;94(3):539-546. [doi: 10.1007/s00420-020-01600-y] [Medline: 33184687]
63. Mitchell DC, Castro J, Armitage TL, Tancredi DJ, Bennett DH, Schenker MB. Physical activity and common tasks of California farm workers: California Heat Illness Prevention Study (CHIPS). J Occup Environ Hyg 2018 Dec;15(12):857-869. [doi: 10.1080/15459624.2018.1519319] [Medline: 30183551]
64. Nazarian N, Liu S, Kohler M, Lee JK, Miller C, Chow WT, et al. Project Coolbit: can your watch predict heat stress and thermal comfort sensation? Environ Res Lett 2021 Feb 23;16(3):034031. [doi: 10.1088/1748-9326/abd130]
65. Notley SR, Akerman AP, Friesen BJ, Poirier MP, Sigal RJ, Flouris AD, et al. Heat tolerance and occupational heat exposure limits in older men with and without type 2 diabetes or hypertension. Med Sci Sports Exerc 2021 Oct 01;53(10):2196-2206. [doi: 10.1249/MSS.0000000000002698] [Medline: 33988544]
66. Ojha A, Shakerian S, Habibnezhad M, Jebelli H, Lee S, Fardhosseini MS. Feasibility of using physiological signals from a wearable biosensor to monitor dehydration of construction workers. In: Proceedings of the Creative Construction e-Conference 2020. 2020 Presented at: CCC '20; June 28-July 1, 2020; Virtual p. 20-28. [doi: 10.3311/ccc2020-004]
67. Pancardo P, Acosta FD, Hernández-Nolasco JA, Wister MA, López-de-Ipiña D. Real-time personalized monitoring to estimate occupational heat stress in ambient assisted working. Sensors (Basel) 2015 Jul 13;15(7):16956-16980 [FREE Full text] [doi: 10.3390/s150716956] [Medline: 26184218]
68. Quante M, Wang R, Weng J, Kaplan ER, Rueschman M, Taveras EM, et al. Seasonal and weather variation of sleep and physical activity in 12-14-year-old children. Behav Sleep Med 2019;17(4):398-410 [FREE Full text] [doi: 10.1080/15402002.2017.1376206] [Medline: 28922020]
69. Raval A, Dutta P, Tiwari A, Ganguly PS, Sathish LM, Mavalankar D, et al. Effects of occupational heat exposure on traffic police workers in Ahmedabad, Gujarat. Indian J Occup Environ Med 2018;22(3):144-151 [FREE Full text] [doi: 10.4103/ijoem.IJOEM_125_18] [Medline: 30647516]
70. Ravanelli NM, Gagnon D, Hodder SG, Havenith G, Jay O. The biophysical and physiological basis for mitigated elevations in heart rate with electric fan use in extreme heat and humidity. Int J Biometeorol 2017 Feb;61(2):313-323. [doi: 10.1007/s00484-016-1213-0] [Medline: 27473105]
71. Ravanelli NM, Hodder SG, Havenith G, Jay O. Heart rate and body temperature responses to extreme heat and humidity with and without electric fans. JAMA 2015 Feb 17;313(7):724-725. [doi: 10.1001/jama.2015.153] [Medline: 25688786]
72. Relf R, Willmott A, Flint MS, Beale L, Maxwell N. Reliability of a wearable sweat rate monitor and routine sweat analysis techniques under heat stress in females. J Therm Biol 2019 Jan;79:209-217. [doi: 10.1016/j.jtherbio.2018.12.019] [Medline: 30612681]
73. Relf R, Eichhorn G, Waldock K, Flint MS, Beale L, Maxwell N. Validity of a wearable sweat rate monitor and routine sweat analysis techniques using heat acclimation. J Therm Biol 2020 May;90:102577. [doi: 10.1016/j.jtherbio.2020.102577] [Medline: 32479383]
74. Rosenthal DG, Vittinghoff E, Tison GH, Pletcher MJ, Olgin JE, Grandis DJ, et al. Assessment of accelerometer-based physical activity during the 2017-2018 California wildfire seasons. JAMA Netw Open 2020 Sep 01;3(9):e2018116 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18116] [Medline: 32997120]
75. Runkle JD, Cui C, Fuhrmann C, Stevens S, Del Pinal J, Sugg MM. Evaluation of wearable sensors for physiologic monitoring of individually experienced temperatures in outdoor workers in southeastern U.S. Environ Int 2019 Aug;129:229-238 [FREE Full text] [doi: 10.1016/j.envint.2019.05.026] [Medline: 31146157]
76. Sugg MM, Fuhrmann CM, Runkle JD. Temporal and spatial variation in personal ambient temperatures for outdoor working populations in the southeastern USA. Int J Biometeorol 2018 Aug;62(8):1521-1534. [doi: 10.1007/s00484-018-1553-z] [Medline: 29802500]
77. Sahu S, Sett M, Kjellstrom T. Heat exposure, cardiovascular stress and work productivity in rice harvesters in India: implications for a climate change future. Ind Health 2013;51(4):424-431 [FREE Full text] [doi: 10.2486/indhealth.2013-0006] [Medline: 23685851]
78. Seo Y, DiLeo T, Powell JB, Kim JH, Roberge RJ, Coca A. Comparison of estimated core body temperature measured with the BioHarness and rectal temperature under several heat stress conditions. J Occup Environ Hyg 2016 Aug;13(8):612-620. [doi: 10.1080/15459624.2016.1161199] [Medline: 26954265]
79. Shakerian S, Habibnezhad M, Ojha A, Lee G, Liu Y, Jebelli H, et al. Assessing occupational risk of heat stress at construction: a worker-centric wearable sensor-based approach. Saf Sci 2021 Oct;142:105395. [doi: 10.1016/j.ssci.2021.105395]
80. Shin M, Swan P, Chow CM. The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. Sleep Sci 2015;8(1):9-15 [FREE Full text] [doi: 10.1016/j.slsci.2015.02.003] [Medline: 26483937]
81. Wang S, Richardson MB, Wu CY, Cholewa CD, Lungu CT, Zaitchik BF, et al. Estimating occupational heat exposure from personal sampling of public works employees in Birmingham, Alabama. J Occup Environ Med 2019 Jun;61(6):518-524 [FREE Full text] [doi: 10.1097/JOM.0000000000001604] [Medline: 30985616]
82. Uejio CK, Morano LH, Jung J, Kintziger K, Jagger M, Chalmers J, et al. Occupational heat exposure among municipal workers. Int Arch Occup Environ Health 2018 Aug;91(6):705-715. [doi: 10.1007/s00420-018-1318-3] [Medline: 29869703]

83. Van Hoye K, Boen F, Lefevre J. Validation of the SenseWear Armband in different ambient temperatures. J Sports Sci 2015;33(10):1007-1018. [doi: 10.1080/02640414.2014.981846] [Medline: 25537112]
84. Williams AA, Spengler JD, Catalano P, Allen JG, Cedeno-Laurent JG. Building vulnerability in a changing climate: indoor temperature exposures and health outcomes in older adults living in public housing during an extreme heat event in Cambridge, MA. Int J Environ Res Public Health 2019 Jul 04;16(13):2373 [FREE Full text] [doi: 10.3390/ijerph16132373] [Medline: 31277359]
85. Xiong J, Lan L, Lian Z, De dear R. Associations of bedroom temperature and ventilation with sleep quality. Sci Technol Built Environ 2020 May 13;26(9):1274-1284. [doi: 10.1080/23744731.2020.1756664]
86. Zheng C, Huang WY, Wong SH. Associations of weather conditions with adolescents' daily physical activity, sedentary time, and sleep duration. Appl Physiol Nutr Metab 2019 Dec;44(12):1339-1344. [doi: 10.1139/apnm-2019-0309] [Medline: 31063697]
87. Zhu ML, Ouyang Q, Shen HG, Zhu YX. Field study on the objective evaluation of sleep quality and sleeping thermal environment in summer. Energy Build 2016 Dec;133:843-852. [doi: 10.1016/j.enbuild.2016.10.007]
88. Kuras ER, Richardson MB, Calkins MM, Ebi KL, Hess JJ, Kintziger KW, et al. Opportunities and challenges for personal heat exposure research. Environ Health Perspect 2017 Aug 01;125(8):085001 [FREE Full text] [doi: 10.1289/EHP556] [Medline: 28796630]
89. Godfrey A, Aranda C, Hussain A, Barreto M, Rocha T, Vitório R. Wearables beyond borders: a case study of barriers to gait assessment in low-resource settings. Maturitas 2020 Jul;137:7-10. [doi: 10.1016/j.maturitas.2020.04.013] [Medline: 32498939]
90. Davies EH, Fieggen K, Wilmshurst J, Anyanwu O, Burman RJ, Komarzynski S. Demonstrating the feasibility of digital health to support pediatric patients in South Africa. Epilepsia Open 2021 Dec;6(4):653-662 [FREE Full text] [doi: 10.1002/epi4.12527] [Medline: 34310860]
91. 'Utoikamanu F. Closing the Technology Gap in Least Developed Countries. United Nations. 2018 Dec. URL: https://www.un.org/en/chronicle/article/closing-technology-gap-least-developed-countries [accessed 2022-03-21]
92. Number of connected wearable devices worldwide by region 2015-2022. Statista. 2022 Feb 14. URL: https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/ [accessed 2022-03-21]
93. Barteit S, Boudo V, Ouedraogo A, Zabré P, Ouremi L, Sié A, et al. Feasibility, acceptability and validation of wearable devices for climate change and health research in the low-resource contexts of Burkina Faso and Kenya: study protocol. PLoS One 2021 Sep 30;16(9):e0257170 [FREE Full text] [doi: 10.1371/journal.pone.0257170] [Medline: 34591893]
94. Sabatini F. Setting up and managing automatic weather stations for remote sites monitoring: from Niger to Nepal. In: Tiepolo M, Pezzoli A, Tarchiani V, editors. Renewing Local Planning to Face Climate Change in the Tropics. Cham, Switzerland: Springer; 2017:21-39.
95. Strigaro D, Cannata M, Antonovic M. Boosting a weather monitoring system in low income economies using open and non-conventional systems: data quality analysis. Sensors (Basel) 2019 Mar 08;19(5):1185 [FREE Full text] [doi: 10.3390/s19051185] [Medline: 30857141]
96. Corona Datenspende. Robert Koch-Institut. URL: https://corona-datenspende.de/science/ [accessed 2021-11-29]
97. Guo Y, Liu X, Peng S, Jiang X, Xu K, Chen C, et al. A review of wearable and unobtrusive sensing technologies for chronic disease management. Comput Biol Med 2021 Feb;129:104163 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.104163] [Medline: 33348217]
98. Hargett T. The Effect of Heat Exposure on Heart Rate and VO2 during Time to Exhaustion Exercise. Kent State University. 2019. URL: https://oaks.kent.edu/ugresearch/2019/effect-heat-exposure-heart-rate-and-vo2-during-time-exhaustion-exercise [accessed 2022-03-10]
99. Okamoto-Mizuno K, Mizuno K. Effects of thermal environment on sleep and circadian rhythm. J Physiol Anthropol 2012 May 31;31(1):14 [FREE Full text] [doi: 10.1186/1880-6805-31-14] [Medline: 22738673]
100. James CA, Richardson AJ, Watt PW, Willmott AG, Gibson OR, Maxwell NS. Short-term heat acclimation improves the determinants of endurance performance and 5-km running performance in the heat. Appl Physiol Nutr Metab 2017 Mar;42(3):285-294. [doi: 10.1139/apnm-2016-0349] [Medline: 28177747]
101. Morrissey MC, Brewer GJ, Williams WJ, Quinn T, Casa DJ. Impact of occupational heat stress on worker productivity and economic cost. Am J Ind Med 2021 Dec;64(12):981-988. [doi: 10.1002/ajim.23297] [Medline: 34590324]
102. Vant-Hull B, Ramamurthy P, Havlik B, Jusino C, Corbin-Mark C, Schuerman M, et al. The Harlem heat project: a unique media–community collaboration to study indoor heat waves. Bull Am Meteorol Soc 2018 Dec 1;99(12):2491-2506. [doi: 10.1175/bams-d-16-0280.1]
103. Mattingly SM, Grover T, Martinez GJ, Aledavood T, Robles-Granda P, Nies K, et al. The effects of seasons and weather on sleep patterns measured through longitudinal multimodal sensing. NPJ Digit Med 2021 Apr 28;4(1):76 [FREE Full text] [doi: 10.1038/s41746-021-00435-2] [Medline: 33911176]
104. Turrisi TB, Bittel KM, West AB, Hojjatinia S, Hojjatinia S, Mama SK, et al. Seasons, weather, and device-measured movement behaviors: a scoping review from 2006 to 2020. Int J Behav Nutr Phys Act 2021 Feb 04;18(1):24 [FREE Full text] [doi: 10.1186/s12966-021-01091-1] [Medline: 33541375]

105. Matsui K. HVAC system for maintaining pleasant sleep during warm nights in Japan. In: Proceedings of the IEEE 30th International Conference on Advanced Information Networking and Applications. 2016 Presented at: AINA '16; March 23-25, 2016; Crans-Montana, Switzerland p. 893-900. [doi: 10.1109/aina.2016.18]
106. Yang H, Cao B, Ju Y, Zhu Y. The effects of local cooling at different torso parts in improving body thermal comfort in hot indoor environments. Energy Build 2019 Sep;198:528-541. [doi: 10.1016/j.enbuild.2019.06.004]

Abbreviations

HR: heart rate
HSI: heat stress index
IET: individually experienced temperature
PICO: Population/Patients, Intervention, Comparison, and Outcome
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
WBGT: wet bulb globe temperature

Edited by L Buis, G Eysenbach; submitted 13.05.22; peer-reviewed by M Raimi, S Bhattacharjee, K Waters; comments to author 15.07.22; revised version received 20.07.22; accepted 26.07.22; published 09.09.22.

Please cite as:
Koch M, Matzke I, Huhn S, Gunga HC, Maggioni MA, Munga S, Obor D, Sié A, Boudo V, Bunker A, Dambach P, Bärnighausen T, Barteit S
Wearables for Measuring Health Effects of Climate Change–Induced Weather Extremes: Scoping Review
JMIR Mhealth Uhealth 2022;10(9):e39532
URL: https://mhealth.jmir.org/2022/9/e39532
doi:10.2196/39532
PMID:36083624

Original Paper

The Intersection of Persuasive System Design and Personalization in Mobile Health: Statistical Evaluation

Aleise McGowan[1], PhD; Scott Sittig[2], RHIA, MHI, PhD; David Bourrie[3], PhD; Ryan Benton[3], PhD; Sriram Iyengar[4], PhD

[1]School of Computing Sciences and Computer Engineering, The University of Southern Mississippi, Hattiesburg, MS, United States

[2]University of Louisiana at Lafayette, Lafayette, LA, United States

[3]University of South Alabama, Mobile, AL, United States

[4]University of Arizona, Pheonix, AZ, United States

Corresponding Author:
Scott Sittig, RHIA, MHI, PhD
University of Louisiana at Lafayette
PO Box 43565
Lafayette, LA, 70504
United States
Phone: 1 3374826160
Email: scott.sittig@louisiana.edu

Abstract

Background: Persuasive technology is an umbrella term that encompasses software (eg, mobile apps) or hardware (eg, smartwatches) designed to influence users to perform preferable behavior once or on a long-term basis. Considering the ubiquitous nature of mobile devices across all socioeconomic groups, user behavior modification thrives under the personalized care that persuasive technology can offer. However, there is no guidance for developing personalized persuasive technologies based on the psychological characteristics of users.

Objective: This study examined the role that psychological characteristics play in interpreted mobile health (mHealth) screen perceived persuasiveness. In addition, this study aims to explore how users' psychological characteristics drive the perceived persuasiveness of digital health technologies in an effort to assist developers and researchers of digital health technologies by creating more engaging solutions.

Methods: An experiment was designed to evaluate how psychological characteristics (self-efficacy, health consciousness, health motivation, and the Big Five personality traits) affect the perceived persuasiveness of digital health technologies, using the persuasive system design framework. Participants (n=262) were recruited by Qualtrics International, Inc, using the web-based survey system of the XM Research Service. This experiment involved a survey-based design with a series of 25 mHealth app screens that featured the use of persuasive principles, with a focus on physical activity. Exploratory factor analysis and linear regression were used to evaluate the multifaceted needs of digital health users based on their psychological characteristics.

Results: The results imply that an individual user's psychological characteristics (self-efficacy, health consciousness, health motivation, and extraversion) affect interpreted mHealth screen perceived persuasiveness, and combinations of persuasive principles and psychological characteristics lead to greater perceived persuasiveness. The *F* test (ie, ANOVA) for model 1 was significant ($F_{9,6540}$=191.806; P<.001), with an adjusted R^2 of 0.208, indicating that the demographic variables explained 20.8% of the variance in perceived persuasiveness. Gender was a significant predictor, with women having higher perceived persuasiveness (P=.008) relative to men. Age was a significant predictor of perceived persuasiveness with individuals aged 40 to 59 years (P<.001) and ≥60 years (P<.001). Model 2 was significant ($F_{13,6536}$=341.035; P<.001), with an adjusted R^2 of 0.403, indicating that the demographic variables self-efficacy, health consciousness, health motivation, and extraversion together explained 40.3% of the variance in perceived persuasiveness.

Conclusions: This study evaluates the role that psychological characteristics play in interpreted mHealth screen perceived persuasiveness. Findings indicate that self-efficacy, health consciousness, health motivation, extraversion, gender, age, and education significantly influence the perceived persuasiveness of digital health technologies. Moreover, this study showed that varying combinations of psychological characteristics and demographic variables affected the perceived persuasiveness of the primary persuasive technology category.

(JMIR Mhealth Uhealth 2022;10(9):e40576) doi:10.2196/40576

KEYWORDS

persuasive technology; personalization; psychological characteristics; self-efficacy; health consciousness; health motivation; personality traits; mobile health; mHealth; mobile phone

Introduction

Background

Given the ubiquitous nature of mobile devices across all socioeconomic groups, digital health technologies have demonstrated their efficacy as key components in educating and treating patients [1]. Mobile health (mHealth) uses mobile devices to practice medicine and public health. Unlike clinic-based treatments, where health care data are sparingly personalized, the ever-present nature of digital health technologies allows for an extensive and more intimate treatment plan. Although digital health technologies allow for the real-time transfer of user data, which allows for more intimate user interaction, these technologies are met with a unique set of challenges, such as creating and maintaining engagement [2]. The efficacy of digital health technologies relies strongly on their ability to continuously engage and re-engage users [3]. The closed-loop engagement process begins with engagement and continuously moves through disengagement to allow the patient to re-engage upon disengagement [4,5]. Properly engaging patients has repeatedly been shown to improve patient outcomes [2].

However, at the core of engagement using digital health technologies, there remains a gap in the literature on how to successfully design these tools based on an individual's dynamic psychological makeup. For instance, there remains a need to learn more about how mHealth treatments work and how to make them more effective. In particular, research on the impact of certain intervention features on user engagement is an important next step in the development of theory and evaluation to develop a science for user engagement [6]. Although the positive influence of persuasion on changing an individual's attitude and behavior has been established [7,8], researchers have contended the need for personalized systems that address individual's personalities to increase the effectiveness of these tools [9,10]. One-size-fits-all digital health technologies that target behavior change to improve the user's health often fail because they do not target the psychological traits that drive an individual's motivations and behaviors, partly because of the lack of guidance from intervention designers and data scientists with numerous options [11]. A dynamic personalized approach to developing persuasive technologies is imperative, as research has shown that strategies that may influence change in an individual with one psychological type may dissuade another individual with a different psychological type [12].

User engagement is a widely used multifaceted term that extends beyond a user's desire to use digital health technologies to the depth of the user's investment [13]. Digital health technologies developers are often tasked with developing tools designed to engage patients, yet little emphasis has been placed on understanding what motivates users to engage with digital health technologies. Developers must move past using a cookie-cutter, one-size-fits-all solution, and seek to develop digital health technologies designed to traverse the fluid terrain that navigates between the expectations of the user and the technological capabilities of the tool. The fluid nature of goals and user preferences determined by user characteristics must also be considered in order to foster various engagement trajectories with digital health technologies. Synonymous with the engagement process, the development of digital health technologies must be dynamic in nature, traversing between design and redesign guided by use [14]. The unconscious disregard for the interdependency among technology, human characteristics, and the socioeconomic environment has been determined to be one of the factors in digital health technologies failing to sustain innovations in the health care field [15,16].

Persuasive technology has emerged as a significant contributor to patient engagement and is used practically in every area of health and wellness [7,17]. Persuasive technology is an umbrella term that encompasses any software (eg, mobile apps) or hardware (eg, smartwatches) designed to influence users to either perform a preferable behavior once or on a long-term basis. These modifications must be achieved without the use of deception, coercion, or inducements [18,19]. By adequately applying persuasive technology, intervention developers have the potential to improve patient outcomes by successfully closing the engagement loop. The modification of user behavior thrives under personalized care that persuasive technology must offer. However, absent from the current literature is adequate information on *how* app designers are to operationalize persuasive design principles based on a more user-centric view [20]. Research is immersed in studies related to the user experience derived from metrics and quantifications, but there remains a void in the literature seeking a more intimate view of the consumer and how they interact with persuasive principles to help guide design processes. The design process is further impaired by a lack of understanding of the psychological characteristics of digital health technology users [21]. Previous research has focused on the development of theories concentrated on predicting acceptance or adherence instead of guiding persuasive technology design principles [22]. This research is needed to fill the gap in the literature addressing the user-centric development of persuasive technologies and developing a better understanding of the psychological characteristics necessary for the successful engagement of digital health technology users.

Consumer and Patient Engagement

There is consensus that an implicit level of engagement is required for digital health technologies to be effective. The absence of engagement impedes digital health technologies from attaining their full potential [23]. This emerging stream of research is built on a somewhat challenging and unstable foundation, as the authors used various procedures to measure engagement [24]. With various metrics in play, the ability to

quantify engagement is a daunting and challenging task [24,25]. This ambiguity further exacerbates our efforts to assess effective engagement.

Digital health technology developers must exercise quantitative and qualitative methods when designing engaging applications [26]. Quantitative measures evaluating intensity and breadth of use are often used to determine the level of consumer engagement [27]. Such a holistic view is not always feasible for developers, but the use of tangible metrics (eg, the amount of screen time of the digital health technology and the number of likes and shares) can be quantified and used for quantitative data [3]. For engagement to be meaningful, digital health technologies must modify user behavior and advance ordinary experiences into aesthetically pleasing ones [28].

Chapman et al [29] proposed that engagement was dichotomous, being either less passive or more passive based on the level of control. More controlled engagement requires information processing such as critical thinking and reasoning and involves a less passive state of engagement. Passive engagement requires less control and is easier to achieve, because the level of effort and motivation is low. Although easier to achieve and maintain, passive engagement is less useful in the successful achievement of established goals that require high levels of cognition [29].

The delivery of appropriately tailored digital health technology content can increase users' engagement and positively influence outcomes. This makes it imperative to understand how to design digital health technologies based on patient and consumer preferences [30]. Identifying the features of digital health technologies that stimulate user engagement is crucial for developing effective tools [21]. One of the key factors in the development of digital health technologies that enhance engagement through the aforementioned techniques is persuasive technology.

Characteristics such as gender, age, and personality affect how users respond to persuasive technologies, causing a pivot from one-size-fits-all solutions to a more user-centric approach [12]. Persuasive technologies can adapt to the individualized characteristics of users, increasing their likelihood of changing their behavior or attitude [31]. Studies show that persuasive technologies that personalize content instead of using one-size-fits-all approaches are more successful in effectively persuading users [32-34]. One-size-fits-all persuasive technologies can be enhanced when a user's individual attitudes and characteristics are used to influence and personalize the persuasiveness of the intervention [35].

Although research has shown that individualized persuasive technology is more effective than persuasive technology designed from a one-size-fits-all perspective [36-38], developers often fail to consider the individualized behavior of stakeholders and how it impacts achieving a target behavior [39]. Digital health technologies that deviate from compartmentalized one-size-fits-all approaches offer a medium through which health care providers can meet the growing demands of users, preferring a more personalized approach [26,40]. This growing demand necessitates the ability to understand how to design digital health technologies that are dynamic enough to accommodate the differing predispositions of end users [30].

Designers must understand how to tailor digital health technologies according to individual characteristics to effectively engage users with these tools. By tailoring digital health technologies to users' characteristics, developers can deliver guidance that is appropriate, relevant, and has a positive impact on engagement [41]. Disregard for the interconnectedness between human characteristics and technology is one reason digital health technologies inevitably become high technology with little to no impact [42]. Current theories are inept at informing digital health technology developers on how to develop and evaluate more adaptive interventions [43,44]. Recognizing the psychological characteristics of end users will allow developers to systematically approach the integration of persuasive design components into digital health technologies.

Data-centered persuasive technologies seek to modify user attitudes or behaviors through users' behavioral data [45]. Current technology allows intervention designers to dynamically generate personalized interventions based on a specific user's personal characteristics [46]. Dynamic approaches acknowledge that interventions designed for one user may not necessarily fit the model required to effectively engage another user. User characteristics often dictate the most effective persuasive technique [35]. Persuasive technologies applicable to the health care domain are more effective when personalized based on the user's personal characteristics [47]. Because personalized persuasive techniques evoke a different response from more traditional, one-size-fits-all techniques, intervention designers must shift to a more individualized approach guided by the individual's preferences [12].

Personalized interventions that target nuances that drive users' choices and behaviors are better suited to facilitate effective engagement than black box, one-size-fits-all solutions [11]. It has long been established that personalized content is more effective as it increases user attention, leading to effective engagement [32]. The application of data collected from individuals is a more advanced method of persuasion that increases the probability of success and results in more active and effective intervention [45]. Determining the key data elements to collect to enhance perceived persuasiveness is critical in efforts to improve engagement (both in the short and long term).

Psychological Characteristics

Self-efficacy

Self-efficacy is loosely defined as an individual's belief that they are capable of successfully executing courses of action required to successfully produce specific behaviors [48]. An individual's estimate of self-efficacy varies in 3 dimensions: magnitude (the individual's belief in their ability to complete a task), strength (the individual's confidence that they are capable of completing various components or varying levels of difficulty in a task), and generality (the extent to which an individual's self-efficacy transfers from one task to related tasks) [48,49]. Self-efficacy is regarded as a core premise of human performance, as demonstrated by its use across multiple domains including education [50,51], exercise [52], physical activity [53], career [54], and health care [55].

Individuals avoid tasks that they presume to exceed their ability levels [56]. Situations in which these tasks occur affect an individual's evaluation of self-efficacy. Self-efficacy is more likely to increase when individuals are able to ascribe success, as opposed to failure, to their individual skill set [56,57]. The difficulty level of a task also correlates with an individual's appraisal of self-efficacy [58]. Tasks deemed difficult to successfully complete tend to have a negative effect on an individual's appraisal of self-efficacy [59]. According to Bandura [56], individuals will go so far as to be unwilling to attempt to manage situations where their low self-efficacy indicates a negative outcome [60].

Hypothesis 1: self-efficacy will positively influence interpreted mHealth screen perceived persuasiveness.

Health Consciousness

Health consciousness is defined as the measure to which an individual integrates health concerns into their daily regime [61-63]. Unlike health motivation (HM), which is external in nature, health consciousness refers to "how" an individual achieves a healthy lifestyle [61]. Research has shown that the higher an individual's health consciousness, the more likely they are to adopt a lifestyle grounded in health behaviors such as fitness and nutritional activities [62,64]. These individuals are cognizant of their health and therefore influenced to adopt these healthier behaviors needed to improve or maintain their health [65].

Studies have shown that health consciousness can positively influence engagement in health-oriented actions [66]. This motivation to engage in health-oriented actions has the propensity to push individuals to become connoisseurs of health information via media sources such as television [64] and the internet [67]. Also observed has been the correlation between the increase in health consciousness and the increase in preventive health care [61,68]. Individuals with high health consciousness reportedly seek to develop and preserve a healthy lifestyle [69].

Hypothesis 2: health consciousness will positively influence interpreted mHealth screen perceived persuasiveness.

Health Motivation

HM is closely related to health consciousness, as it is one of the 3 elements that comprise health consciousness [67]. HM is an individual's drive to engage in health-related activities to improve or maintain preventive health behaviors [61,70]. HM has been found to be a relatively consistent state deeply rooted in an individual's psychological composition [61]. Research has shown that HM serves as the source of an individual's desire, adoption, and practice of preventive health behaviors [61,70]. Motivation has been found to be both competency-based (whether a person can achieve the goal) and goal-oriented (the way a task is managed is determined by the individual's objective) [71].

It has also been determined that HM can gauge an individual's well-being with regard to health behavior–related concerns and actions [72] and drive consumer engagement in health maintenance behaviors [70]. HM is directly linked to an individual's internal characteristics [61]. Research has consistently shown that internalized motivation results in more pronounced adherence to preventive health behaviors such as weight loss [73,74]. Whether an individual expects to succeed also plays a key role in their degree of motivation [75].

Hypothesis 3: HM will positively influence interpreted mHealth screen perceived persuasiveness.

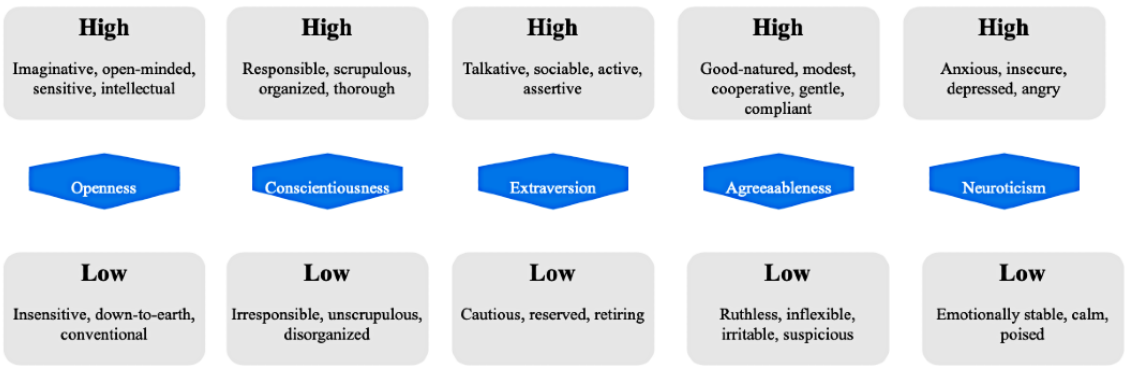
Personality Traits

Personality traits and strategies used to engage users have an impact on the effective engagement of digital health technology [76]. Understanding these personality traits is critical for creating digital health solutions that meet the needs of users. One of the most commonly used personality models is the Big Five factor model [77]. The Big Five factor framework was developed by Goldberg [78] and later validated by Costa and McCrae [78,79]. This model delineated five factors of personality:

1. Openness to experience: the extent to which an individual requires intellectual stimulation, change, and variety
2. Conscientiousness: the extent to which an individual is willing to comply with conventional rules, norms, and standards
3. Extraversion: the extent to which an individual needs attention and social interaction
4. Agreeableness: the extent to which an individual needs pleasant and harmonious relationships with others
5. Neuroticism: the extent to which an individual observes the world as threatening and beyond their control [80]

Each Big Five personality category can be regarded as a continuum in which individual scores range from high to low (Figure 1 [77]).

Figure 1. Big Five continuum.



Frequent time constraints are the drivers for a more succinct measurement tool [81]. The Mini-International Personality Item Pool (Mini-IPIP) Scale is a condensed 20-item diagnostic tool that has been validated in multiple studies [82]. Researchers have used the Big Five framework to predict user characteristics across a conglomerate of domains: career [83], relationship satisfaction and love styles [84], academic performance [85], preventive health care [86], and more.

Individual personality traits often reflect not only what drives and motivates people but also what they prefer. The Big Five personality dimensions describe human behavior in 5 dimensions: openness, conscientiousness, extraversion or introversion, agreeableness or disagreeableness, and neuroticism. Individual personality traits should be an antecedent of consumer engagement with mHealth apps [81].

Hypothesis 4: openness will positively influence interpreted mHealth screen perceived persuasiveness.

Hypothesis 5: conscientiousness will positively influence interpreted mHealth screen perceived persuasiveness.

Hypothesis 6: extraversion will positively influence interpreted mHealth screen perceived persuasiveness.

Hypothesis 7: agreeableness will positively influence interpreted mHealth screen perceived persuasiveness.

Hypothesis 8: neuroticism will negatively influence interpreted mHealth screen perceived persuasiveness.

Methods

Ethics Approval

Institutional review board approval was obtained from the University of South Alabama (application 18-353/1314060-1).

Overview

To examine the factors related to engagement behavior with the intention to use an mHealth app, a multiple-phase experiment was conducted in the summer of 2020. This experiment involved a survey-based design with a series of 25 mHealth app screens that featured the use of persuasive principles, with a focus on physical activity. This study used exploratory factor analysis (EFA) and multiple linear regression to aid designers in the user-centric development of persuasive technologies. This study aimed to develop a better understanding of the psychological characteristics necessary for the successful engagement of digital health technology users.

Recruitment

Participants were recruited by Qualtrics International, Inc to use the web-based survey system by XM Research Service [87], which has been previously used by researchers in a variety of disciplines [88,89]. Qualtrics reimbursed participants with a predetermined amount of money arranged between Qualtrics and participants. Once interested participants were selected by Qualtrics, they were directed to the informed consent page via an anonymous link. Upon consenting to participate, they were directed to a web-based engagement screen survey. Participants were recruited between July 23, 2020, and August 3, 2020. The engagement screen survey took an average of 28.08 (19.35 SD) minutes to complete. There were 273 completed survey responses; however, 11 (4%) were deleted owing to evident signs of respondents being “speeders” that completed the survey in an impossibly quick time or “straight lining” and giving identical answer choices repeatedly, leaving this study with 262 (95.9%) viable responses.

Screen Development

To examine the perceived persuasiveness of mHealth screens, 25 unique mHealth screens were developed following the persuasive system design (PSD) categories and principles developed by Oinas-Kukkonen and Harjumaa [90]. The screens were all developed with the central theme of improving or increasing exercise as a use case.

The mHealth screen development process began with the creation of a wireframe prototype [91]. The prototype was created on sheets of paper, with each sheet representing one of the mHealth app screens. The initial step for each prototype was to document the persuasive system category, design principle, and targeted implementation as per Oinas-Kukkonen [90]. A brief description of the details of the screen was then added to the prototype, followed by the mHealth screen being given a reference name based on the details in the write-up used throughout the questionnaire development and analysis process. Table 1 presents examples of the initial steps. A sketch of the prototype was then drawn based on the documentation so that each sheet would represent one of the mHealth screens.

BuildFire [92] was used to develop a digital high-fidelity prototype for each mobile app screen. These prototypes were used to support the design goals established in the initial prototype. Once the prototypes were developed, iPhone XS Max was used to create still images of the mHealth screens using the screenshot function. This method was used so that the image would visually represent what a user would see on their smartphone. The images were exported from the mobile phone to a laptop computer via email. Once the prototypes were exported, 2 experts in the field of persuasive technology conducted a blind review to validate the mHealth screen, representing the persuasive technology principle intended by the author. The expert review panel consisted of a reviewer with 12 years of experience in the field of persuasive technology and a reviewer with 9 years of experience in the field of persuasive technology. Each expert created a datasheet with the associated screen names and listed the PSD principles identified on each screen.

Following the expert inspection and blind review, a consultation was held with the expert review panel, where notes and suggestions were reviewed. The review and modification processes continued until the developer and reviewers reached a consensus. The mHealth screens were iteratively evaluated, modified, and improved following each expert inspection and blind review. For the initial round, 23 mHealth screens were developed: Add, Start, Burpee-Squat, Increase, Mountain, Target, Trophy, Late, Calories, Dinner Chat, Tracker, About Us, Stories, Leaderboard, Journal, Partners, Ads, Strategy, CDC, HIPAA, Contact, Before After, and Yoga. During the initial expert review, the developer and reviewers identified 11

mHealth screens with conflicting persuasive technology principles that required modification: Target, Dinner Chat, About Us, Journal, Partners, Strategy, HIPAA, Contact, Before After, Yoga, and CDC. CDC was dropped following the initial review because the designed persuasive category was not seen by either of the 2 reviewers, and the category that was identified was seen on another screen. The Apple mHealth screen was created to replace the CDC and submitted with revisions for round 2. Consensus was reached on the 23 mHealth screens during the second round. In addition, 3 paper and high-fidelity prototypes were created for the remaining mHealth screens (SSL, Avatar, and Recreation) following the aforementioned methods. Additional mHealth screens were iteratively evaluated, modified, and improved using expert inspection and blind review methods used during rounds 1 and 2. The iterative process resulted in 25 mHealth screens designed for the questionnaire that were agreed upon through the blind review process, and an mHealth screen prototype was discarded. The acceptance of the mHealth screen by round is presented in Table 2.

Table 1. Examples of the initial prototype development steps.

Persuasive system category and design principle	Targeted implementation	Mock-up	Mock-up name
Primary task support			
Reduction	Provide simple steps for an activity	Show literature such as weight loss made simple, which gives simple steps to get started for losing weight	Start
Tunneling	Guiding people in a process step by step to meet a goal	Fitness program with step-by-step workout plan. Once daily or weekly goals are reached, the next set of steps are given	Burpee-Squat
Tailoring	The system uses factors relevant to the individual to motivate the users based on their needs, interests, personality, and so on	Users can modify the app to reflect their interests and personality (change color pallet, select what is displayed on home screen, etc)	Add
Personalization	Suggestions, praise, and rewards are given at appropriate time to motivate users to stay on track	Increase the user's activity goal based on accomplishments or modify dietary plan based on weight loss	Increase
Self-monitoring	Allows users to follow or monitor their performance to ensure that they are staying on track	Summary of daily or weekly activity calculations and weekly weight summaries	Tracker
System credibility support			
Trustworthiness	Apps should appear to be truthful, fair, and unbiased	Display information guaranteeing HIPAA[a] compliance to reassure users that information will not be shared with third-party organizations	HIPAA
Expertise	Provide content from experts (physicians or specialists)	Chat screen showing interaction with person that resembles a physician or medical professional	About Us

[a]HIPAA: Health Insurance Portability and Accountability Act.

Table 2. Mobile health (mHealth) screen acceptance by round.

Screen name	Round 1	Round 2	Round 3
Add	✓[a]		
Start	✓		
Burpee-Squat	✓		
Increase	✓		
Mountain	✓		
Target		✓	
Trophy	✓		
Late	✓		
Calories	✓		
Dinner Chat		✓	
Tracker	✓		
About Us		✓	
Stories	✓		
Leaderboard	✓		
Journal		✓	
Partners		✓	
Ads	✓		
Strategy		✓	
CDC[b]	Dropped	N/A[c]	N/A
HIPAA[d]		✓	
Contact		✓	
Before After		✓	
Yoga		✓	
Apple	N/A	Replaced CDC	
SSL	N/A	N/A	✓
Avatar	N/A	N/A	✓
Recreation	N/A	N/A	✓

[a]✓: indicates that the mHealth screen was accepted.

[b]CDC: Centers for Disease Control and Prevention.

[c]N/A: not applicable.

[d]HIPAA: Health Insurance Portability and Accountability Act.

The primary task support category aids the user in performing fundamental tasks by reducing complex tasks into simpler tasks. The primary task principles include reduction, tunneling, tailoring, personalization, self-monitoring, simulation, and rehearsal [90]. Textbox 1 describes the primary task support design principles [90].

The dialogue support category facilitates human-to-computer dialogue between the persuasive system and user. The principles used to provide feedback are praise, rewards, reminders, suggestions, similarities, liking, and social roles [90]. Textbox 2 describes the principles of the dialogue support category [90].

The system credibility category represents how systems can be made more persuasive by making them more credible. The principles used to give credibility include trustworthiness, expertise, surface credibility, real-world feel, authority, third-party endorsements, and verifiability [90]. Textbox 3 describes the principles of the system credibility category [90].

Principles in the social support category motivate systems through social influence. Design principles in this category include social facilitation, social comparison, normative influence, social learning, cooperation, competition, and recognition. Textbox 4 shows the principles of social support [90].

Table 3 depicts the final iteration of testing and includes the principles per screen and the principle category. Table 4 shows the percentage of screens in the primary persuasive technology

category. Figure 2 shows one of the final mHealth screens developed. A visual representation of all 25 screens is available in Multimedia Appendix 1. Of the 25 screens developed, 5 (20%) screens had a primary principle from the primary task support category, 7 (28%) had a primary principle from the dialogue support category, 8 (32%) had a primary principle from the system credibility support category, and 5 (20%) had a primary principle from the social support category.

Textbox 1. Primary task principles.

Persuasive system category, design principle, and principle description—primary task support

- Reduction
  - Provides simple steps for an activity
- Tunneling
  - Guides people in a process step by step to meet a goal
- Tailoring
  - Uses factors relevant to the individual to motivate the users based on their needs, interests, personality, and so on
- Personalization
  - Suggestions, praise, and rewards are given at appropriate time to motivate users to stay on track
- Self-monitoring
  - Allows users to follow or monitor their performance to ensure they are staying on track
- Simulation
  - Allows the user to observe the cause-and-effect link regarding their behavior
- Rehearsal
  - Allows users to rehearse a behavior

Textbox 2. Dialogue support principles.

Persuasive system category, design principle, and principle description—dialogue support

- Praise
  - Uses images, words, sounds, and so on to praise the user for their behavior
- Rewards
  - Uses web-based rewards, given to the user for performing tasks related to the target behavior
- Reminders
  - Reminds the user of their target behavior
- Suggestion
  - Offers the user suggestions that fit the target behavior
- Similarity
  - Remind users of themselves in some way
- Liking
  - The digital health technology should be visually attractive
- Social role
  - The digital health technology adopts a social role

Textbox 3. System credibility principles.

Persuasive system category, design principle, and principle description—system credibility support

- Trustworthiness
  - Apps should appear to be truthful, fair, and unbiased
- Expertise
  - Provide content from sources that are knowledgeable and competent
- Surface credibility
  - Systems should visually appear to be competent and credible
- Real-world feel
  - Systems should highlight the people or organizations that are providing content by providing information about them
- Authority
  - Systems should leverage roles of authority by referring to organizations and people that are seen as authority figures
- Third-party endorsements
  - Systems should provide users with endorsements from third parties that are well known and trusted
- Verifiability
  - Systems should provide ways for users to easily use external sources to verify the accuracy of the content

Textbox 4. Social support principles.

Persuasive system category, design principle, and principle description—social support

- Social learning
  - The digital health technology should target behavior by providing the user with a way to observe other users who are performing the same target behavior
- Social comparison
  - The digital health technology should motivate the user by allowing them to compare their performance with other users who are performing the same task
- Normative influence
  - The digital health technology should use normative influence or peer pressure
- Social facilitation
  - The digital health technology should allow users to perceive that other users are using the system to perform the target behavior along with them
- Cooperation
  - The digital health technology should leverage the users' natural drive to cooperate
- Competition
  - The digital health technology should leverage the users' natural drive to compete with other users
- Recognition
  - The digital health technology should offer users public recognition

Table 3. Mobile app screen name with persuasive principles and categories.

Screen name	Principle 1 (primary)	Principle 2	Principle 3
Add	PT[a]: tailoring	PT: tunneling	—[b]
Start	PT: reduction	PT: tunneling	—
Burpee-Squat	PT: tunneling	PT: reduction	—
Increase	DS[c]: praise	—	—
Mountain	PT: rehearsal	DS: suggestion	—
Target	DS: praise	PT: personalization	—
Trophy	DS: rewards	DS: praise	—
Late	DS: reminders	—	—
Calories	DS: suggestion	—	—
Dinner Chat	DS: social role	DS: praise	—
Tracker	PT: self-monitoring	—	—
About Us	SC[d]: expertise	SC: trustworthiness	SC: authority
Stories	SS[e]: recognition	PT: simulation	DS: praise
Leaderboard	SS: competition	—	—
Journal	SS: social learning	SS: social comparison	SC: social facilitation
Partners	SC: trustworthiness	SC: expertise	SC: authority
Ads	SC: surface credibility	—	—
Strategy	SC: authority	SC: expertise	—
Apple	SC: verifiability	SC: expertise	SC: authority
HIPAA[f]	SC: trustworthiness	SC: surface credibility	—
Contact	SC: real-world feel	—	—
Before After	SC: normative influence	PT: simulation	—
Yoga	SS: cooperation	DS: praise	SS: social comparison
SSL	SC: third-party endorsements	SC: trustworthiness	—
Avatar	DS: similarity	DS: liking	—

[a]PT: primary task support.
[b]Not available.
[c]DS: dialogue support.
[d]SC: system credibility support.
[e]SS: social support.
[f]HIPAA: Health Insurance Portability and Accountability Act.

Table 4. Screen category breakdown.

Persuasive technology category	Mobile screens (%)
Primary task support	20
Dialogue support	28
System credibility support	32
Social support	20

Figure 2. Sample mobile health screen developed and accepted during review.

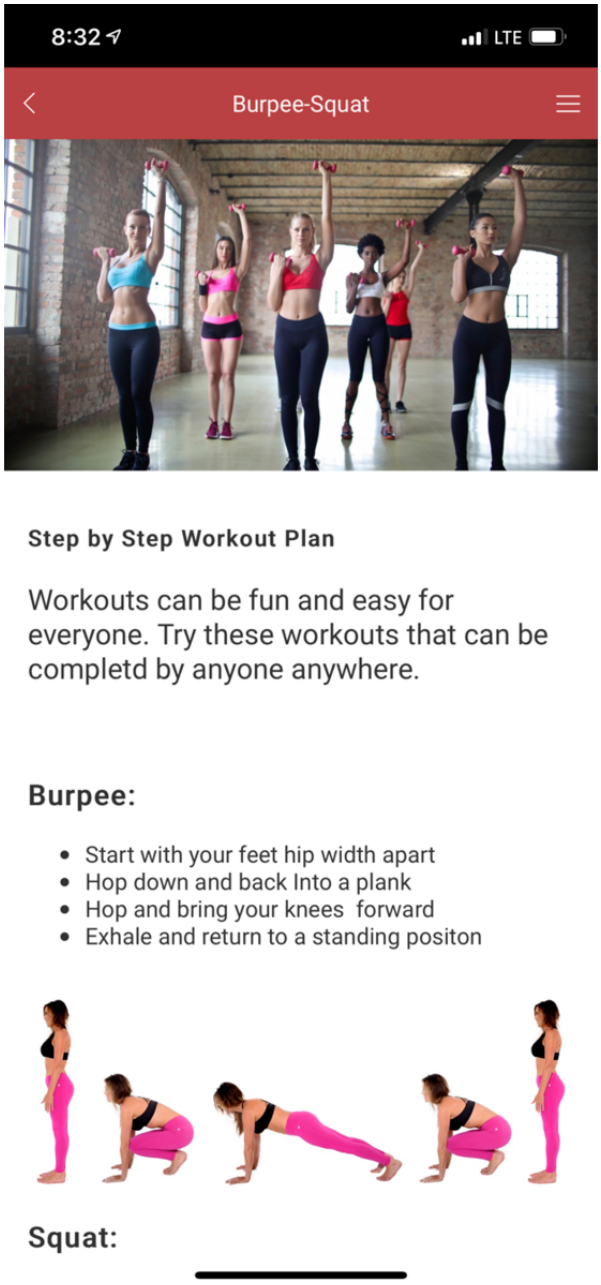
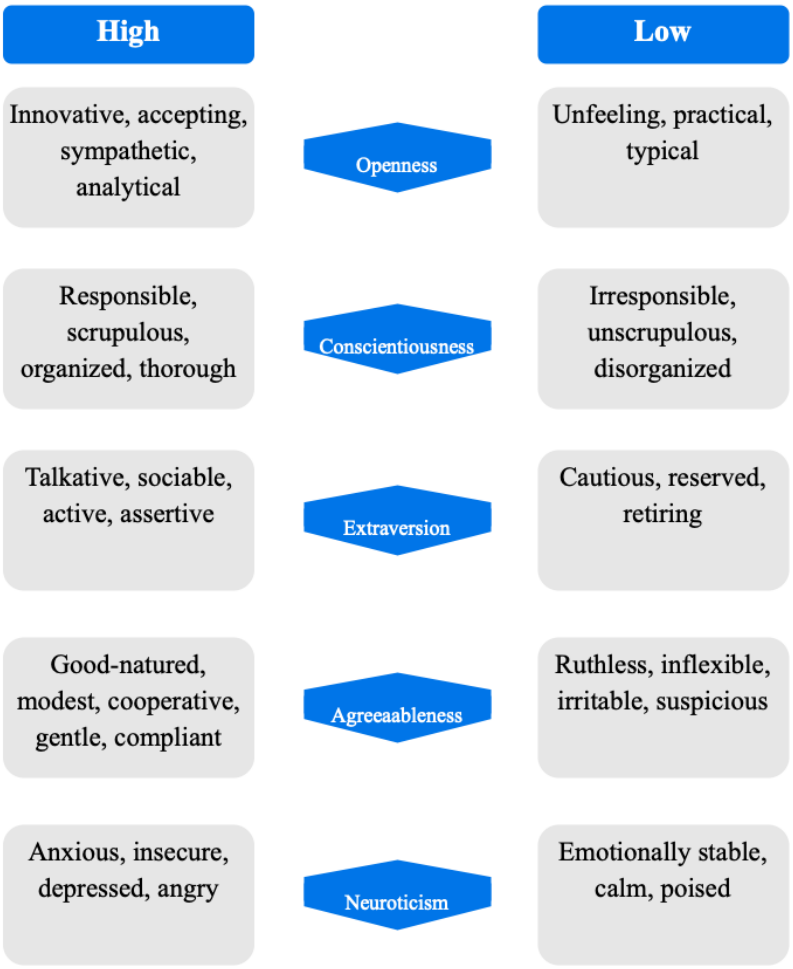


Figure 5. Big Five continuum.



Measurement Items

New General Self-Efficacy Scale

After completion of the social demographic information, the participants were asked 8 questions about their self-efficacy using the New General Self-Efficacy Scale (Multimedia Appendix 2) by Chen et al [93]. The work by Chen et al [93] extends the work by Bandura [48,56], which focuses on the magnitude and strength dimensions of self-efficacy and includes the generality dimension of self-efficacy. Data on self-efficacy were collected from participants at baseline and 20 days after the first survey. The 7-item Likert scale used in this study ranged from strongly disagree to strongly agree.

Health Consciousness

Participants were then asked to complete 6 questions about their health consciousness using the Health Consciousness Scale by Jayanti and Burns [61] (Multimedia Appendix 3), which was adapted from the original Health Consciousness Scale by Kraft and Goodell [64]. The development of the health consciousness scale was facilitated by borrowing items from the literature to generate items for scales. Multiple items were used to measure each of the constructs proposed, with purification steps taken during the development of the scales. The 7-item Likert scale used in this study ranged from strongly disagree to strongly agree. Types of health consciousness questions the participants encountered include “I am interested in information about my health” and “I read more health-related articles than I did 3 years ago.”

Health Motivation

Participants were then asked to complete questions about their HM using the Health Motivation Scale by Jayanti and Burns [61] (Multimedia Appendix 4). Scale development was facilitated by borrowing items from the literature, and generating items for scales was used to develop the HM scale. The scale development and purification followed well-established procedures reported in the literature. This section consists of 6 questions using a 7-point Likert scale ranging from strongly disagree to strongly agree. Participants answered questions about their HM, such as, “I try to prevent common health problems before I feel any symptoms” and “I would rather enjoy life than try to make sure I am not exposing myself to health risks.”

Personality Traits

Finally, participants were asked to answer personality questions that generally described them as they were now and not as they wished to be in the future. The participants completed the Mini-IPIP Scale by Donnellan et al [82] (Multimedia Appendix 5), which consists of 20 questions focusing on extraversion, agreeableness, conscientiousness, neuroticism, and intellect. The stability of the Mini-IPIP Scale was measured at multiple intervals. The initial study was conducted at intervals of a few

weeks, and the subsequent study was conducted over several months. The questions were answered using a 7-point Likert scale ranging from extremely inaccurate to extremely accurate. Participants rated the accuracy of statements, such as, “Am the life of the party” and “Am not really interested in others.”

Perceived Persuasiveness

After answering the psychological questions, participants were asked to complete questions about the perceived persuasiveness of the individual mHealth screens. The participants completed the Perceived Persuasiveness Scale by Lehto et al [94] (Multimedia Appendix 6), which consists of 3 questions using a 7-point Likert scale ranging from strongly disagree to strongly agree. During the development of the scale, data examining perceived persuasiveness were collected from participants at baseline and 2 and 6 weeks after the intervention. During the study, participants answered questions, at the screen level, about the perceived persuasiveness of the mHealth app screens, such as, “This mobile health screen has an influence on me” and “This mobile health screen makes me reconsider my overall health and wellness.”

Results

Exploratory Factor Analysis

EFA was conducted using SPSS to appraise the factor structure of the survey items. More specifically, principal component factoring using a Promax rotation was the extraction method for this analysis [95]. Kaiser normalization (eigenvalue>1) was used to determine the number of extracted factors. As the factor-loading cutoff varies in the literature, this research used a conventional liberal-to-conservative continuum, with all factor loadings of ≥0.4 being considered salient for this study, and cross-loadings >0.2 were considered for elimination [96,97].

The initial iteration of the EFA was conducted on 8 self-efficacy items, 6 health consciousness items, 6 HM items, 20 Big Five items, 3 perceived persuasiveness items, 3 intention items, 4 willingness to use items, and 4 marker variable questions (n=62). A total of 16 items (HM_1, HM_2, Big Five-Conscientiousness (R) Q8, Big Five-Conscientiousness (R) Q18, all 4 Big Five Agreeableness items, all 4 Big Five Openness items, and all 4 Big Five Neuroticism items were eliminated owing to cross-loading issues. In addition, 2 items—Big Five-Conscientiousness Q3 and Big Five-Conscientiousness Q13—were eliminated for having correlation coefficients below the threshold and failing to load properly on other items.

For the final stage, principal component factor analysis of the remaining 29 items resulted in 6 extracted 6 factors explaining 73.67% of the variance. The factor-loading matrix for the final solution is presented in Table 5. Hypotheses 4, 5, 7, and 8 were untestable because of the EFA results. Marker variables were removed from further statistical analyses after a lack of correlation was confirmed through EFA analysis.

Table 5. Final exploratory factor analysis results.

	Factor					
	1	2	3	4	5	6
SE[a] Q1	0.736	—[b]	—	—	—	—
SE Q2	0.872	—	—	—	—	—
SE Q3	0.902	—	—	—	—	—
SE Q4	0.908	—	—	—	—	—
SE Q5	0.914	—	—	—	—	—
SE Q6	0.793	—	—	—	—	—
SE Q7	0.686	—	—	—	—	—
SE Q8	0.821	—	—	—	—	—
HC[c] Q1	—	0.817	—	—	—	—
HC Q2	—	0.848	—	—	—	—
HC Q3	—	0.782	—	—	—	—
HC Q4	—	0.714	—	—	—	—
HC Q5	—	0.653	—	—	—	—
HC Q6	—	0.457	—	—	—	—
HM[d] Q3	—	—	0.781	—	—	—
HM Q4	—	—	0.847	—	—	—
HM Q5	—	—	0.878	—	—	—
HM Q6	—	—	0.728	—	—	—
TF_PP[e] Q1	—	—	—	0.973	—	—
TF_PP Q2	—	—	—	0.999	—	—
TF_PP Q3	—	—	—	0.989	—	—
MV[f] 1	—	—	—	—	0.858	—
MV 2	—	—	—	—	0.821	—
MV 3	—	—	—	—	0.710	—
MV 4	—	—	—	—	0.830	—
E[g] Q1	—	—	—	—	—	0.408
E Q6	—	—	—	—	—	0.624
E Q11	—	—	—	—	—	0.566
E Q16	—	—	—	—	—	0.768

[a]SE: self-efficacy.

[b]Not available.

[c]HC: health consciousness.

[d]HM: health motivation.

[e]TF_PP: perceived persuasiveness.

[f]MV: marker variable.

[g]E: extraversion.

Statistical Results

Weighted scores were computed for self-efficacy, health consciousness, HM, extraversion, and perceived persuasiveness, using the final EFA factor loadings. Table 6 presents the Cronbach α, mean, SD, and intercorrelation among the variables included in this study.

Linear regression analysis was performed for weighted variables. A total of 2 linear regression models were used. Table 7 shows the regression coefficients. Model 1 included the demographic control variables of gender, age, and education level as

predictors of perceived persuasiveness. The demographic variables were dummy coded with "male," "under 40," and "less than high school" as the reference category for gender, age, and education level. The *F* test (ie, ANOVA) for model 1 was significant ($F_{9,6540}$=191.806; *P*<.001), with an adjusted R^2 of 0.208, indicating that the demographic variables explained 20.8% of the variance in perceived persuasiveness. Gender was a significant predictor, with women having higher perceived persuasiveness (B=0.127, SE=0.048; t_{6540}=2.668; *P*=.008) and nonbinary individuals having lower perceived persuasiveness (B=−2.856, SE=0.265; t_{6540}=−10.767; *P*<.001) relative to men. Age was a significant predictor, with individuals aged 40 to 59 age years (B=−0.643, SE=0.069; t_{6540}=−9.377; *P*<.001) and ≥60 years (B=−2.116, SE=0.059; t_{6540}=−35.752; *P*<.001) having lower perceived persuasiveness relative to individuals aged ≤40 years. Education level was a significant predictor, as individuals who held associate degrees (B=−0.411, SE=0.163; t_{6540}=−2.514; *P*=.01) and Bachelor's degrees (B=−0.581, SE=0.157; t_{6540}=−3.696; *P*<.001) tended to have lower perceived persuasiveness than individuals who had not completed high school.

In model 2, the theorized effects were added as predictors. The *F* test for model 2 was significant ($F_{13,6536}$=341.035; *P*<.001), with an adjusted R^2 of 0.403, indicating that the demographic variables self-efficacy, health consciousness, HM, and extraversion together explained 40.3% of the variance in perceived persuasiveness. Table 7 presents the regression coefficients for model 2. The nonbinary category of sex remained a significant predictor; however, the female sex category was no longer significant in model 2 (B=−0.002, SE=0.042; t_{6536}=−0.048; *P*=.96). Both age categories were significant predictors in model 2. The associate and Bachelor's categories of education level remained significant predictors in model 2, and the categories of some college (B=−0.462, SE=0.137; t_{6536}=−3.378; *P*<.001) and graduate degree (B=−0.555, SE=0.139; t_{6536}=−3.985; *P*<.001) became significant in model 2. Self-efficacy was a significant positive predictor (B=0.263, SE=0.026; t_{6536}=10.174; *P*<.001), indicating that individuals with higher self-efficacy tended to have higher perceived persuasiveness. Health consciousness was a significant positive predictor (B=0.883, SE=0.022; t_{6536}=40.000; *P*<.001), indicating that individuals with higher health consciousness tended to have higher perceived persuasiveness. HM was a significant positive predictor (B=0.200, SE=0.017; t_{6536}=11.597; *P*<.001), indicating that individuals with higher HM tended to have higher perceived persuasiveness. Extraversion was a significant positive predictor (B=0.150, SE=0.026; t_{6536}=5.884; *P*<.001), indicating that individuals with higher extraversion tended to have higher perceived persuasiveness. The results of significant hypothesis testing are summarized in Table 8.

Table 6. Correlation matrix for weighted variables[a].

Variable	Mean (SD)	Cronbach α	1	2	3	4	5	6
Self-efficacy	4.574 (0.852)	.939	.833	—[b]	—	—	—	—
Health consciousness	3.455 (0.994)	.858	.239[c]	.724	—	—	—	—
Health motivation	3.071 (1.205)	.862	.067[c]	−0.132[c]	.811	—	—	—
Extraversion	2.110 (0.816)	.699	.263[c]	.142[c]	−0.069[c]	.605	—	—
Perceived persuasiveness	3.822 (2.047)	.977	.283[c]	.529[c]	.081[c]	.159[c]	.987	—
Marker variable	3.089 (1.135)	.840	.153[c]	.391[c]	.303[c]	−0.030[d]	.363[c]	.807

[a]Values on the diagonal are the square roots of the average variance extracted.

[b]Not available.

[c]*P*<.01.

[d]*P*<.05.

Table 7. Results for multiple linear regression modelsa (N=6550).

Variable	Model 1[a]			Model 2[b]		
	B (SE)	*t* test (*df*)	Significance (*P* value)	B (SE)	*t* test (*df*)	Significance (*P* value)
Constant	5.406 (0.161)	33.531 (6540)	<.001	0.005 (0.202)	0.023 (6537)	.98
Control variables						
Gender (female)	0.127 (0.048)	2.668 (6540)	.008	−0.002 (0.042)	−0.048 (6537)	.96
Gender (nonbinary)	−2.856 (0.265)	−10.767 (6540)	<.001	−2.239 (0.238)	−9.412 (6537)	<.001
Age (40-59 years)	−0.643 (0.069)	−9.377 (6540)	<.001	−0.477 (0.061)	−7.869 (6537)	<.001
Age (≥60 years)	−2.116 (0.059)	−35.752 (6540)	<.001	−1.388 (0.054)	−25.816 (6537)	<.001
Education (high school graduate)	−0.302 (0.161)	−1.880 (6540)	.06	−0.218 (0.140)	−1.555 (6537)	.120
Education (some college, no degree)	−0.279 (0.156)	−1.782 (6540)	.08	−0.462 (0.137)	−3.378 (6537)	<.001
Education (associate degree)	−0.411 (0.163)	−2.514 (6540)	.01	−0.389 (0.142)	−2.731 (6537)	.006
Education (Bachelor's degree)	−0.581 (0.157)	−3.696 (6540)	<.001	−0.624 (0.137)	−4.542 (6537)	<.001
Education (graduate degree)	−0.059 (0.159)	−0.370 (6540)	.71	−0.555 (0.139)	−3.985 (6537)	<.001
Theorized effects						
Self-efficacy	—[c]	—	—	0.263 (0.026)	10.174 (6537)	<.001
Health consciousness	—	—	—	0.883 (0.022)	40.000 (6537)	<.001
Health motivation	—	—	—	0.200 (0.017)	11.597 (6537)	<.001
Extraversion	—	—	—	0.150 (0.026)	5.884 (6537)	<.001

[a]Model 1: R^2=0.208

[b]Model 2: R^2=0.403.

[c]—: indicates that the theorized effects weren't added until model 2.

Table 8. Results of tested hypotheses.

Hypothesis	Result
Hypothesis 1: self-efficacy will positively influence interpreted mHealth screen perceived persuasiveness.	Supported
Hypothesis 2: health consciousness will positively influence interpreted mHealth screen perceived persuasiveness.	Supported
Hypothesis 3: health motivation will positively influence interpreted mHealth screen perceived persuasiveness.	Supported
Hypothesis 6: extraversion will positively influence interpreted mHealth screen perceived persuasiveness.	Supported

Discussion

Principal Findings

To the best of our knowledge, this study is the first to use a combination of self-efficacy, health consciousness, HM, extraversion, gender, age, and education to examine their impact on the effective engagement of users of digital health technologies. By integrating psychological characteristics, this study advances the current understanding of how psychological characteristics affect the perceived persuasiveness of persuasive technology. To evaluate this, the researchers examined the impact of psychological characteristics (self-efficacy, health consciousness, HM, and Big Five personality traits) on the perceived persuasiveness of digital health technologies. Using the PSD framework, this study was designed to evaluate how these psychological characteristics affect the perceived persuasiveness of digital health technologies. In addition, the dynamic intertwining of psychological characteristics that drives the perceived persuasiveness of the primary PSD technique categories was illuminated through multiple linear regression analysis.

Furthermore, this study opens a pathway for designers of digital health technologies to gain further knowledge on why individual characteristics must be considered during the design process. Keizer et al [42] suggested that misalignment between end users and digital technologies is often a result of developers failing to consider the end user during the development process. Although the benefits of personalizing persuasive systems have been acknowledged, the field is still in its infancy, and there is very little knowledge on the best way to tailor these technologies [98,99]. The findings from this study suggest that using a

dynamic, data-centered approach that considers that the end users' self-efficacy, health consciousness, HM, extraversion, age, gender, and education could be a way to increase the perceived persuasiveness of digital health technologies.

In addition, this research offers developers vital information pertaining to user-centric development of persuasive digital health technologies. The information gained can be used by designers to increase the perceived persuasiveness of digital health technologies by providing guidance on how to dynamically use PSD principles based on an individual's psychological characteristics and demographic makeup. These PSD principles can be delivered in various components such as virtual reality or health care gaming approaches, which can further establish a stronger connection to an individual's psychological characteristics [100].

On the basis of these major findings, the role of self-efficacy should be considered by persuasive technology designers. Statistical analysis found self-efficacy to be a significant positive predictor of perceived persuasiveness. Multiple linear regression analyses found that health consciousness was a significant positive predictor of perceived persuasiveness. In addition, the model found HM to be a significant positive predictor of perceived persuasiveness. Multiple linear regression analyses also found extraversion to be a significantly positive predictor of perceived persuasiveness. These findings are important, as they shed additional light on which psychological characteristics influence a user's perceived persuasiveness. In addition, it helps validate why one-size-fits-all approaches do not necessarily work. The findings suggest that individuals with low self-efficacy and low health consciousness will not necessarily be influenced (perceived persuasiveness) by the same mHealth app design as those with higher self-efficacy and health consciousness levels.

Demographic data, such as age and gender, should also be considered by developers of digital health technologies. The findings strongly suggest that the distribution of perceived persuasiveness shifts from negatively skewed to positively skewed as an individual ages. In addition, this shift occurs earlier in women (ie, aged 40-59 years) than in men who do not shift until the oldest age group (ie, aged ≥60 years). The perceived persuasiveness by age group and gender is available in Multimedia Appendix 7. This was an interesting and unexpected finding, and additional research is required. Potentially, these findings can represent the aging process for which health consciousness, for example, has increased owing to typical chronic diseases that manifest as individuals age.

Future Research and Limitations

Despite the theoretical and practical contributions of this study, there are limitations to the generalizability of the findings. Further examination of the demographic data showed that only 7.3% (19/262) of participants were between the ages of 18 and 29 years. Additional research should be conducted that focuses on the younger population, aged 18 to 29 years.

The research only examined extraversion due to multicollinearity issues with other items from the Big Five personality traits. Sleep et al [101] found that longer measures contain considerably more variance than shorter, more condensed measures. Further studies should use a more extensive Big Five personality test such as the Neo Personality Inventory [102] rather than the Mini-IPIP Scale [82].

The Adult Hope Scale by Snyder et al [103] was also dropped from the study owing to multicollinearity issues with the New General Self-Efficacy Scale by Chen et al [93]. It was observed that all the self-efficacy constructs and adult hope constructs were cross-loading; therefore, adult hope was eliminated because self-efficacy is regarded as a core premise of human performance across multiple domains, and adult hope measurements conceptually and operationally function synonymously as self-efficacy [104].

Multicollinearity issues were also identified among perceived persuasiveness, intention, and willingness to use; therefore, the intention and willingness to use constructs were eliminated from the model because perceived persuasiveness was studied across multiple domains, and perceived persuasiveness was more pursuant to this study.

A key limitation of this study is the use of static screens. A fully developed app will allow researchers to evaluate the engagement of digital health tools. Running these studies in tandem will allow researchers to evaluate engagement on both sides to see if higher perceived persuasiveness leads to higher engagement.

Conclusions

This study aimed to examine how users' psychological characteristics influence the perceived persuasiveness of digital health technologies. This research contributes to advancing the field of data-driven, user-centric development of persuasive technologies by investigating the intertwining of users' psychological characteristics and the perceived persuasiveness of digital health technologies. This work opens a new research avenue by examining the role of psychological characteristics in interpreting the perceived persuasiveness of mHealth screens. The use of dynamic data-driven capabilities is important for advancing perceived persuasiveness, which has the potential to engage users of digital health technologies successfully.

This work also describes the roles that psychological characteristics play in interpreting mHealth screen perceived persuasiveness. Evidence has shown that self-efficacy, health consciousness, HM, extraversion, gender, age, and education significantly influence the perceived persuasiveness of digital health technologies. Moreover, this study showed that varying combinations of psychological characteristics and demographic variables affected the perceived persuasiveness of the primary persuasive technology category. Incorporating these psychological characteristics and demographic variables should allow digital health technology developers to overcome the gap stemming from one-size-fits-all approaches.

On the basis of the findings of this research, mHealth app researchers and developers should design apps that dynamically interact with users using psychological characteristics and demographics to drive the persuasive techniques presented to the user. This process should include a pre-enrollment assessment, for which the user's psychological characteristics are evaluated before deployment of the mHealth app. This would

allow for the right persuasive techniques to be deployed in an attempt to better engage the user, which can potentially lead to more favorable behavior. Moving from a “one-size-fits-all” to a personalized persuasive approach has the potential to create long-term engagement, which has plagued mHealth researchers and developers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Developed mobile health screens.
[PDF File (Adobe PDF File), 1205 KB - mhealth_v10i9e40576_app1.pdf]

Multimedia Appendix 2
New General Self-Efficacy Scale.
[DOCX File , 16 KB - mhealth_v10i9e40576_app2.docx]

Multimedia Appendix 3
Health Consciousness Scale.
[DOCX File , 16 KB - mhealth_v10i9e40576_app3.docx]

Multimedia Appendix 4
Health Motivation Scale.
[DOCX File , 16 KB - mhealth_v10i9e40576_app4.docx]

Multimedia Appendix 5
Big Five Mini-International Personality Item Pool Scale.
[DOCX File , 17 KB - mhealth_v10i9e40576_app5.docx]

Multimedia Appendix 6
Perceived Persuasiveness Scale.
[DOCX File , 15 KB - mhealth_v10i9e40576_app6.docx]

Multimedia Appendix 7
Perceived persuasiveness by age and gender.
[PDF File (Adobe PDF File), 128 KB - mhealth_v10i9e40576_app7.pdf]

References

1. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
2. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
3. O'Brien H. A holistic approach to measuring user engagement. In: Filimowicz M, Tzankova V, editors. New Directions in Third Wave Human-Computer Interaction: Volume 2 - Methodologies. Cham, Switzerland: Springer; 2018:81-102.
4. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong KL, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
5. 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]
6. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
7. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91 [FREE Full text] [doi: 10.1177/1460458216650979] [Medline: 27245673]
8. Silva M, Graham F, Levack W, Hay-Smith J. Persuasive technology and behaviour change in parent-focused eHealth interventions supporting child health: a scoping review protocol. N Z J Physiother 2019 Mar 31;47(1):36-48 [FREE Full text] [doi: 10.15619/NZJP/47.1.05]

9. Wall HJ, Campbell CC, Kaye LK, Levy A, Bhullar N. Personality profiles and persuasion: an exploratory study investigating the role of the Big-5, type D personality and the Dark Triad on susceptibility to persuasion. Pers Individ Dif 2019 Mar;139:69-76. [doi: 10.1016/j.paid.2018.11.003]
10. Kaptein M, Markopoulos P, de Ruyter B, Aarts E. Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. Int J Human Comput Stud 2015 May;77:38-51. [doi: 10.1016/j.ijhcs.2015.01.004]
11. Engl E, Smittenaar P, Sgaier SK. Identifying population segments for effective intervention design and targeting using unsupervised machine learning: an end-to-end guide. Gates Open Res 2019 Oct 21;3:1503 [FREE Full text] [doi: 10.12688/gatesopenres.13029.2] [Medline: 31701090]
12. Abdullahi AM, Oyibo K, Orji R, Kawu AA. The influence of age, gender, and cognitive ability on the susceptibility to persuasive strategies. Information 2019 Nov 15;10(11):352. [doi: 10.3390/info10110352]
13. O'Brien HL, Arguello J, Capra R. An empirical study of interest, task complexity, and search behaviour on user engagement. Inf Process Manag 2020 May;57(3):102226. [doi: 10.1016/j.ipm.2020.102226]
14. Goldkuhl G. From ensemble view to ensemble artefact – an inquiry on conceptualisations of the IT artefact. Syst Sign Action 2013;7(1):49-72.
15. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
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 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
17. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
18. Orji R, Oyibo K, Lomotey RK, Orji FA. Socially-driven persuasive health intervention design: competition, social comparison, and cooperation. Health Informatics J 2019 Dec;25(4):1451-1484 [FREE Full text] [doi: 10.1177/1460458218766570] [Medline: 29801426]
19. Iyengar MS, Florez-Arango JF, Garcia CA. GuideView: a system for developing structured, multimodal, multi-platform persuasive applications. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09; April 26-29, 2009; Claremont, CA, USA p. 1-5. [doi: 10.1145/1541948.1541990]
20. Thomson C, Nash J, Maeder A. Persuasive design for behaviour change apps: issues for designers. In: Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists. 2016 Presented at: SAICSIT '16; September 26-28, 2016; Johannesburg, South Africa p. 1-10. [doi: 10.1145/2987491.2987535]
21. Tuman M, Moyer A. Health intentions and behaviors of health app owners: a cross-sectional study. Psychol Health Med 2019 Aug;24(7):819-826. [doi: 10.1080/13548506.2019.1576911] [Medline: 30729803]
22. Al-Ramahi M, El-Gayar O, Liu J. Discovering design principles for persuasive systems: a grounded theory and text mining approach. In: Proceedings of the 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS '16; January 5-8, 2016; Koloa, HI, USA p. A-83. [doi: 10.1109/hicss.2016.387]
23. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
24. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: a survey-based quantitative study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13244 [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
25. Zagalo N. From experience to engagement. In: Zagalo N, editor. Engagement Design: Designing for Interaction Motivations. Cham, Switzerland: Springer; 2020:11-30.
26. Sahin C. Rules of engagement in mobile health: what does mobile health bring to research and theory? Contemp Nurse 2018;54(4-5):374-387. [doi: 10.1080/10376178.2018.1448290] [Medline: 29502472]
27. Helsper EJ, Eynon R. Distinct skill pathways to digital engagement. Eur J Commun 2013 Sep 16;28(6):696-713. [doi: 10.1177/0267323113499113]
28. Salehzadeh Niksirat K, Sarcar S, Sun H, Law EL, Clemmensen T, Bardzell J, et al. Approaching engagement towards human-engaged computing. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA '18; April 21-26, 2018; Montreal, Canada p. 1-4. [doi: 10.1145/3170427.3185364]
29. Chapman P, Selvarajah S, Webster J. Engagement in multimedia training systems. In: Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999 Presented at: HICSS '99; January 5-8, 1999; Maui, HI, USA. [doi: 10.1109/hicss.1999.772808]
30. Tarute A, Nikou S, Gatautis R. Mobile application driven consumer engagement. Telemat Inform 2017 Jul;34(4):145-156. [doi: 10.1016/j.tele.2017.01.006]
31. Wiafe I. The role of U-FADE in selecting persuasive system features. In: Khosrow-Pour M, editor. Encyclopedia of Information Science and Technology. 4th edition. Hershey, PA, USA: IGI Global; 2018:7785-7795.

32. Gena C, Grillo P, Lieto A, Mattutino C, Vernero F. When personalization is not an option: an in-the-wild study on persuasive news recommendation. Information 2019 Sep 26;10(10):300. [doi: 10.3390/info10100300]
33. Orji R, Mandryk RL, Vassileva J. Gender, age, and responsiveness to Cialdini's persuasion strategies. In: Proceedings of the 10th International Conference on Persuasive Technology. 2015 Presented at: PERSUASIVE '15; June 3-5, 2015; Chicago, IL, USA p. 147-159. [doi: 10.1007/978-3-319-20306-5_14]
34. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User Adap Interact 2014 Jul 14;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
35. Berkovsky S, Freyne J, Oinas-Kukkonen H. Influencing individually: fusing personalization and persuasion. ACM Trans Interact Intell Syst 2012 Jun 01;2(2):1-8. [doi: 10.1145/2209310.2209312]
36. Orji FA, Oyibo K, Orji R, Greer J, Vassileva J. Personalization of persuasive technology in higher education. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. 2019 Presented at: UMAP '19; June 9-12, 2019; Larnaca, Cyprus p. 336-340. [doi: 10.1145/3320435.3320478]
37. Ruijten PA. The similarity-attraction paradigm in persuasive technology: effects of system and user personality on evaluations and persuasiveness of an interactive system. Behav Inf Technol 2021;40(8):734-746. [doi: 10.1080/0144929x.2020.1723701]
38. Orji RO, Vassileva J, Mandryk RL. Modeling gender differences in healthy eating determinants for persuasive intervention design. In: Proceedings of the 8th International Conference on Persuasive Technology. 2013 Presented at: PERSUASIVE '13; April 3-5, 2013; Sydney, Australia p. 161-173. [doi: 10.1007/978-3-642-37157-8_20]
39. Taype GE, Calani MC. Extending persuasive system design frameworks: an exploratory study. In: Proceedings of Information Technology and Systems. 2020 Presented at: ICITS '20; February 5-7, 2020; Bogota, Colombia p. 35-45. [doi: 10.1007/978-3-030-40690-5_4]
40. Almunawar MN, Anshari M, Younis MZ. Incorporating customer empowerment in mobile health. Health Policy Technol 2015 Dec;4(4):312-319. [doi: 10.1016/j.hlpt.2015.08.008]
41. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]
42. Keizer J, Beerlage-de Jong N, al Naiemi N, van Gemert-Pijnen LJ. Persuading from the start: participatory development of sustainable persuasive data-driven technologies in healthcare. In: Proceedings of the 15th International Conference on Persuasive Technology. 2020 Presented at: PERSUASIVE '20; April 20–23, 2020; Aalborg, Denmark p. 113-125. [doi: 10.1007/978-3-030-45712-9_9]
43. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
44. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
45. Shin Y, Kim J. Data-centered persuasion: nudging user's prosocial behavior and designing social innovation. Comput Human Behav 2018 Mar;80:168-178. [doi: 10.1016/j.chb.2017.11.009]
46. Dalecke S, Karlsen R. Designing dynamic and personalized nudges. In: Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics. 2020 Presented at: WIMS '20; June 30-July 3, 2020; Biarritz, France p. 139-148. [doi: 10.1145/3405962.3405975]
47. Abdullahi AM, Orji R, Kawu AA. Gender, age and subjective well-being: towards personalized persuasive health interventions. Information 2019 Sep 27;10(10):301. [doi: 10.3390/info10100301]
48. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY, USA: W. H. Freeman; 1997.
49. Bong M. Generality of academic self-efficacy judgments: evidence of hierarchical relations. J Educ Psychol 1997 Dec;89(4):696-709. [doi: 10.1037/0022-0663.89.4.696]
50. van Dinther M, Dochy F, Segers M, Braeken J. The construct validity and predictive validity of a self-efficacy measure for student teachers in competence-based education. Stud Educ Eval 2013 Sep;39(3):169-179. [doi: 10.1016/j.stueduc.2013.05.001]
51. Bulfone G, Badolamenti S, Biagioli V, Maurici M, Macale L, Sili A, et al. Nursing students' academic self-efficacy: a longitudinal analysis of academic self-efficacy changes and predictive variables over time. J Adv Nurs 2021 May;77(5):2353-2362. [doi: 10.1111/jan.14771] [Medline: 33559919]
52. Simonavice EM, Wiggins MS. Exercise barriers, self-efficacy, and stages of change. Percept Mot Skills 2008 Dec;107(3):946-950. [doi: 10.2466/pms.107.3.946-950] [Medline: 19235423]
53. Koring M, Richert J, Lippke S, Parschau L, Reuter T, Schwarzer R. Synergistic effects of planning and self-efficacy on physical activity. Health Educ Behav 2012 Apr;39(2):152-158. [doi: 10.1177/1090198111417621] [Medline: 22167316]
54. Falco LD, Summers JJ. Improving career decision self-efficacy and STEM self-efficacy in high school girls: evaluation of an intervention. J Career Dev 2017 Jul 24;46(1):62-76. [doi: 10.1177/0894845317721651]
55. Messina R, Rucci P, Sturt J, Mancini T, Fantini MP. Assessing self-efficacy in type 2 diabetes management: validation of the Italian version of the Diabetes Management Self-Efficacy Scale (IT-DMSES). Health Qual Life Outcomes 2018 Apr 23;16(1):71 [FREE Full text] [doi: 10.1186/s12955-018-0901-3] [Medline: 29685153]

56. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
57. Medrano LA, Flores-Kanter E, Moretti L, Pereno GL. Effects of induction of positive and negative emotional states on academic self-efficacy beliefs in college students. Psicología Educativa 2016 Dec;22(2):135-141. [doi: 10.1016/j.pse.2015.03.003]
58. Buckworth J. Promoting self-efficacy for healthy behaviors. ACSM'S Health Fitness J 2017;21(5):40-42. [doi: 10.1249/fit.0000000000000318]
59. Lindenmeier J. Promoting volunteerism: effects of self-efficacy, advertisement-induced emotional arousal, perceived costs of volunteering, and message framing. Voluntas 2008 Mar 15;19(1):43-65. [doi: 10.1007/s11266-008-9054-z]
60. Wigal JK, Creer TL, Kotses H. The COPD self-efficacy scale. Chest 1991 May;99(5):1193-1196. [doi: 10.1378/chest.99.5.1193] [Medline: 2019177]
61. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. J Acad Market Sci 1998 Jan 01;26(1):6. [doi: 10.1177/0092070398261002]
62. Yan M, Filieri R, Raguseo E, Gorton M. Mobile apps for healthy living: factors influencing continuance intention for health apps. Technol Forecast Soc Change 2021 May;166(C):120644. [doi: 10.1016/j.techfore.2021.120644]
63. Chen MF, Lin NP. Incorporation of health consciousness into the technology readiness and acceptance model to predict app download and usage intentions. Internet Res 2018 Apr 04;28(2):351-373. [doi: 10.1108/intr-03-2017-0099]
64. Kraft FB, Goodell PW. Identifying the health conscious consumer. J Health Care Mark 1993;13(3):18-25. [Medline: 10129812]
65. Barauskaite D, Gineikiene J, Fennis BM, Auruskeviciene V, Yamaguchi M, Kondo N. Eating healthy to impress: how conspicuous consumption, perceived self-control motivation, and descriptive normative influence determine functional food choices. Appetite 2018 Dec 01;131:59-67. [doi: 10.1016/j.appet.2018.08.015] [Medline: 30114492]
66. Parashar S, Mungra Y, Sood G. Health consciousness as an enabler for exploratory buying behavior among consumers. SCMS J Indian Manag 2019;16(2):87-102.
67. Ahadzadeh AS, Pahlevan Sharif S, Sim Ong F. Online health information seeking among women: the moderating role of health consciousness. Online Inf Rev 2018 Feb 12;42(1):58-72. [doi: 10.1108/oir-02-2016-0066]
68. Donalds C, Osei-Bryson KM. Cybersecurity compliance behavior: exploring the influences of individual decision style and other antecedents. Int J Inf Manag 2020 Apr;51(C):102056. [doi: 10.1016/j.ijinfomgt.2019.102056]
69. Park J, Ahn J, Yoo WS. The effects of price and health consciousness and satisfaction on the medical tourism experience. J Healthc Manag 2017;62(6):405-417. [doi: 10.1097/JHM-D-16-00016] [Medline: 29135765]
70. Tanner EC, Vann RJ, Kizilova E. Consumer-level perceived access to health services and its effects on vulnerability and health outcomes. J Public Policy Market 2020 Feb 27;39(2):240-255. [doi: 10.1177/0743915620903299]
71. Toste JR, Didion L, Peng P, Filderman MJ, McClelland AM. A meta-analytic review of the relations between motivation and reading achievement for K–12 students. Rev Educ Res 2020 May 13;90(3):420-456. [doi: 10.3102/0034654320919352]
72. Dehghani M, Kim KJ, Dangelico RM. Will smartwatches last? Factors contributing to intention to keep using smart wearable technology. Telemat Inform 2018 May;35(2):480-490 [FREE Full text] [doi: 10.1016/j.tele.2018.01.007]
73. Ferron JC, Elbogen EB, Swanson JW, Swartz MS, McHugo GJ. A conceptually based scale to measure consumers' treatment motivation. Res Soc Work Pract 2010 Jan 05;21(1):98-105. [doi: 10.1177/1049731509357629]
74. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020 Apr;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
75. Muenks K, Wigfield A, Eccles JS. I can do this! The development and calibration of children's expectations for success and competence beliefs. Dev Rev 2018 Jun;48:24-39. [doi: 10.1016/j.dr.2018.04.001]
76. Wagner 3rd B, Liu E, Shaw SD, Iakovlev G, Zhou L, Harrington C, et al. e wrapper: operationalizing engagement strategies in mHealth. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:790-798 [FREE Full text] [doi: 10.1145/3123024.3125612] [Medline: 29362728]
77. Roccas S, Sagiv L, Schwartz SH, Knafo A. The Big Five personality factors and personal values. Pers Soc Psychol Bull 2002 Jun 1;28(6):789-801. [doi: 10.1177/0146167202289008]
78. Goldberg LR. An alternative "description of personality": the big-five factor structure. J Pers Soc Psychol 1990 Dec;59(6):1216-1229. [doi: 10.1037//0022-3514.59.6.1216] [Medline: 2283588]
79. Costa Jr PT, McCrae RR. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Odessa, FL, USA: Psychological Assessment Resources; 1992.
80. Borghans L, Duckworth AL, Heckman JJ, ter Weel B. The economics and psychology of personality traits. J Human Resources 2008;43(4):972-1059. [doi: 10.1353/jhr.2008.0017]
81. Gosling SD, Rentfrow PJ, Swann Jr WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
82. Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. Psychol Assess 2006 Jun;18(2):192-203. [doi: 10.1037/1040-3590.18.2.192] [Medline: 16768595]

83. Wilmot MP, Wanberg CR, Kammeyer-Mueller JD, Ones DS. Extraversion advantages at work: a quantitative review and synthesis of the meta-analytic evidence. J Appl Psychol 2019 Dec;104(12):1447-1470. [doi: 10.1037/apl0000415] [Medline: 31120263]
84. White JK, Hendrick SS, Hendrick C. Big five personality variables and relationship constructs. Pers Individ Dif 2004 Nov;37(7):1519-1530. [doi: 10.1016/j.paid.2004.02.019]
85. Stajkovic AD, Bandura A, Locke EA, Lee D, Sergent K. Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: a meta-analytic path-analysis. Pers Individ Dif 2018 Jan;120:238-245. [doi: 10.1016/j.paid.2017.08.014]
86. Nolan A, McCrory C, Moore P. Personality and preventive healthcare utilisation: evidence from the Irish Longitudinal Study on Ageing. Prev Med 2019 Mar;120:107-112. [doi: 10.1016/j.ypmed.2018.12.029] [Medline: 30660708]
87. Qualtrics. Provo, UT, USA: Qualtrics; 2019. URL: https://www.qualtrics.com/au/ [accessed 2019-07-18]
88. Pangbourne K, Bennett S, Baker A. Persuasion profiles to promote pedestrianism: effective targeting of active travel messages. Travel Behav Soc 2020 Jul;20:300-312. [doi: 10.1016/j.tbs.2020.04.004]
89. Kniffin KM, Bogan VL, Just DR. "Big men" in the office: the gender-specific influence of weight upon persuasiveness. PLoS One 2019 Nov 11;14(11):e0222761 [FREE Full text] [doi: 10.1371/journal.pone.0222761] [Medline: 31710625]
90. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 May 1;24:28. [doi: 10.17705/1CAIS.02428]
91. Rosenzweig E. Successful User Experience: Strategies and Roadmaps. Amsterdam, The Netherlands: Elsevier; 2015.
92. Buildfire Corporation. 2019. URL: https://buildfire.com/ [accessed 2018-10-23]
93. Chen G, Gully SM, Eden D. Validation of a new general self-efficacy scale. Organ Res Method 2001 Jan 1;4(1):62-83. [doi: 10.1177/109442810141004]
94. Lehto T, Oinas-Kukkonen H, Drozd F. Factors affecting perceived persuasiveness of a behavior change support system. In: Proceedings of the 33rd International Conference on Information Systems. 2012 Presented at: ICIS '12; December 16-19, 2012; Orlando, FL, USA p. 1-15 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.959.1225&rep=rep1&type=pdf
95. Williams B, Onsman A, Brown T. Exploratory factor analysis: a five-step guide for novices. Australas J Paramed 2010 Aug 02;8(3):990399. [doi: 10.33151/ajp.8.3.93]
96. Knekta E, Runyon C, Eddy S. One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sci Educ 2019 Mar;18(1):rm1 [FREE Full text] [doi: 10.1187/cbe.18-04-0064] [Medline: 30821600]
97. Matsunaga M. How to factor-analyze your data right: do's, don'ts, and how-to's. Int J Psychol Res 2010 Jun 30;3(1):97-110. [doi: 10.21500/20112084.854]
98. Orji R, Kaptein M, Ham J, Oyibo K, Nwokeji J. Personalizing persuasive technologies: a road map to the future. In: Proceedings of the 13th International Conference on Persuasive Technology. 2018 Presented at: PERSUASIVE '18; April 18-19, 2018; Waterloo, Canada p. 1-4. [doi: 10.1145/3213586.3225246]
99. Oyibo K, Orji R, Ham J, Nwokeji J, Ciocarlan A. Personalizing persuasive technologies: personalization for wellbeing. In: Personalizing Persuasive Technology Workshop. 2019 Presented at: PPT '19; April 9-11, 2019; Limassol, Cyprus. [doi: 10.13140/RG.2.2.26605.00485]
100. Horsham C, Dutton-Regester K, Antrobus J, Goldston A, Price H, Ford H, et al. A virtual reality game to change sun protection behavior and prevent cancer: user-centered design approach. JMIR Serious Games 2021 Mar 25;9(1):e24652 [FREE Full text] [doi: 10.2196/24652] [Medline: 33764308]
101. Sleep CE, Lynam DR, Miller JD. A comparison of the validity of very brief measures of the Big Five/Five-Factor Model of personality. Assessment 2021 Apr;28(3):739-758. [doi: 10.1177/1073191120939160] [Medline: 32762351]
102. McCrae RR, Costa Jr PT, Martin TA. The NEO-PI-3: a more readable revised NEO Personality Inventory. J Pers Assess 2005 Jun;84(3):261-270. [doi: 10.1207/s15327752jpa8403_05] [Medline: 15907162]
103. Snyder CR, Harris C, Anderson JR, Holleran SA, Irving LM, Sigmon ST. Adult Hope Scale (AHS). J Pers Soc Psychol 1991;60:570-585. [doi: 10.1037/t00088-000]
104. Zhou M, Kam CC. Hope and general self-efficacy: two measures of the same construct? J Psychol 2016 Jul 03;150(5):543-559. [doi: 10.1080/00223980.2015.1113495] [Medline: 26761605]

Abbreviations

EFA: exploratory factor analysis
HM: health motivation
mHealth: mobile health
Mini-IPIP: Mini-International Personality Item Pool
PSD: persuasive system design

Edited by L Buis; submitted 27.06.22; peer-reviewed by I Schiering, P Knapp, M Alrige; comments to author 21.07.22; revised version received 04.08.22; accepted 10.08.22; published 14.09.22.

Please cite as:
McGowan A, Sittig S, Bourrie D, Benton R, Iyengar S
The Intersection of Persuasive System Design and Personalization in Mobile Health: Statistical Evaluation
JMIR Mhealth Uhealth 2022;10(9):e40576
URL: https://mhealth.jmir.org/2022/9/e40576
doi:10.2196/40576
PMID:36103226

Original Paper

Mobile for Mothers mHealth Intervention to Augment Maternal Health Awareness and Behavior of Pregnant Women in Tribal Societies: Randomized Quasi-Controlled Study

Avishek Choudhury[1*], PhD; Murari Choudhury[2*], MBA

[1]Industrial and Management Systems Engineering, Benjamin M Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

[2]Network for Enterprise Enhancement and Development Support, Deoghar, India

[*]all authors contributed equally

Corresponding Author:
Avishek Choudhury, PhD
Industrial and Management Systems Engineering
Benjamin M Statler College of Engineering and Mineral Resources
West Virginia University
1306 Evansdale Drive
PO Box 6107
Morgantown, WV, 26506-6107
United States
Phone: 1 3042939431
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: Despite several initiatives taken by government bodies, disparities in maternal health have been noticeable across India's socioeconomic gradient due to poor health awareness.

Objective: The aim of this study was to implement an easy-to-use mobile health (mHealth) app—Mobile for Mothers (MfM)—as a supporting tool to improve (1) maternal health awareness and (2) maternal health–related behavioral changes among tribal and rural communities in India.

Methods: Pregnant women, aged 18 to 45 years, were selected from two rural villages of Jharkhand, India: (1) the intervention group received government-mandated maternal care through an mHealth app and (2) the control group received the same government-mandated care via traditional means (ie, verbally). A total of 800 accredited social health activists (ASHAs) were involved, of which 400 were allocated to the intervention group. ASHAs used the MfM app to engage with pregnant women during each home visit in the intervention group. The mHealth intervention commenced soon after the baseline survey was completed in February 2014. The end-line data were collected between November 2015 and January 2016. We calculated descriptive statistics related to demographics and the percentage changes for each variable between baseline and end line per group. The baseline preintervention groups were compared to the end-line postintervention groups using Pearson chi-square analyses. Mantel-Haenszel tests for conditional independence were conducted to determine if the pre- to postintervention differences in the intervention group were significantly different from those in the control group.

Results: Awareness regarding the five cleans (5Cs) in the intervention group increased (P<.001) from 143 (baseline) to 555 (end line) out of 740 participants. Awareness about tetanus vaccine injections and the fact that pregnant women should receive two shots of tetanus vaccine in the intervention group significantly increased (P<.001) from 73 out of 740 participants (baseline) to 372 out of 555 participants (end line). In the intervention group, awareness regarding the fact that problems like painful or burning urination and itchy genitals during pregnancy are indicative of a reproductive tract infection increased (P<.001) from 15 (baseline) to 608 (end line) out of 740 participants. Similarly, knowledge about HIV testing increased (P<.001) from 39 (baseline) to 572 (end line) out of 740 participants. We also noted that the number of pregnant women in the intervention group who consumed the prescribed dosage of iron tablets increased (P<.001) from 193 (baseline) out of 288 participants to 612 (end line) out of 663 participants.

Conclusions: mHealth interventions can augment awareness of, and persistence in, recommended maternal health behaviors among tribal communities in Jharkhand, India. In addition, mHealth could act as an educational tool to help tribal societies break away from their traditional beliefs about maternal health and take up modern health care recommendations.

Trial Registration: OSF Registries 9U8D5; https://doi.org/10.17605/OSF.IO/9U8D5

(JMIR Mhealth Uhealth 2022;10(9):e38368) doi:10.2196/38368

KEYWORDS

maternal health; mHealth; digital divide; disparity; socioeconomic; India; health; awareness; mobile; intervention; adherence; health behaviors; tribal; community; education

Introduction

Maternal mortality remains a serious health issue for India, and the problem for deprived parts of the population is worse. Government bodies have taken several initiatives [1], but disparities in maternal health services, provision of maternal care, and health outcomes have been noticeable across India's socioeconomic gradient [2-4]. For example, in Kerala, the maternal mortality rate was as low as 61 per 100,000 live births in 2013, whereas in the northern states of Bihar and Jharkhand, it was 208 per 100,000 live births [5]. Interventions to minimize maternal deaths have been primarily noted to be ineffective within the rural population [6]. India's remote parts are populous, and a lack of budget limits the success of providing adequate resources, such as health educators, infrastructure, and time to educate the targeted population [7].

Maternal mortality in rural societies of India, according to the "3 delays" model [8,9], primarily occurs due to three significant factors: (1) delay in deciding to seek care, (2) delay in obtaining timely care, and (3) delay in receiving appropriate treatment. The "3 delays" model adopts a systemic approach reflecting various obligations to avoid maternal mortality at the family, community, and health system levels. Several extrinsic factors often instigate these three delays, including insufficient support for maternal health services and the lack of awareness regarding maternal health care among pregnant women [10].

Over the years, the lack of maternal health awareness has caused several health concerns, including anemia, neural tube defects [11], tetanus infection, immunodeficiency syndrome, and even perinatal deaths [12]. Another problem faced by the tribal communities is that the people in rural India do not make use of the available health care facilities, primarily due to their belief systems. According to a previous study, approximately 21.2% of tribal women in Odisha, India, neglected their critical health conditions and sought home remedies as their primary treatment choice [13]. Certain cultures hold the extreme beliefs that their ailments, including maternal health complications, have been caused by taboos [14].

Mobile health (mHealth) technology has the potential to address the aforementioned maternal health–related concerns by augmenting maternal health awareness. Nowadays, health care is being revolutionized by mHealth. mHealth apps may have educational benefits for patients, as these technologies can offer user-friendly and easy-to-understand personalized treatment or ailment-related education [15]. Overall, mHealth for patient care can be helpful for two fundamental reasons: (1) its ability to resolve systemic obstacles (ie, language and geographical locations) and address global public health concerns [16] and (2) its promise to encourage the patient to be persistent with taking their medication [17].

Thus far, most research in this domain has been conducted in developed nations [18] and with consumers who are educated and typically familiar with mobile phones. The benefits of mHealth interventions are that they can achieve near-complete coverage of the target population. However, a large population of pregnant women in rural regions of India have no access to mobile phones [19] or any mobile technologies in general. Currently, mHealth interventions do not have sufficient evidence showing their impact on maternal health awareness, particularly in tribal and rural communities [20]. Further research may contribute to understanding the impacts of mHealth on maternal health awareness and behaviors in rural and tribal regions of India. Therefore, to address this gap, the Network for Enterprise Enhancement and Development Support (NEEDS), a nongovernmental organization (NGO) in India, and Simavi, a Dutch NGO, conceptualized and developed an easy-to-use mHealth app—Mobile for Mothers (MfM)—as a supporting tool to improve (1) maternal health awareness and (2) maternal health–related behavioral changes among tribal communities. The MfM intervention was conducted in partnership with the Rural Health Mission of the Government of Jharkhand under the European Union–funded Initiative for Transparency and Good Governance.

Methods

Overview

This study was part of a more extensive study conducted in collaboration with the Rural Health Mission of the Government of Jharkhand under the European Union–funded Initiative for Transparency and Good Governance. The findings from this larger study have been reported elsewhere [21-23].

Study Design

This study was conducted with the tribal communities in rural areas of Jharkhand, India. This was a randomized quasi-controlled analysis of two groups (ie, two independent rural villages in Jharkhand, India): (1) the intervention group (village A) received maternal care via government-mandated programs through the mHealth app and (2) the control group (village B) received the same care via government-mandated programs but through traditional means (ie, verbally).

Under the government programs provided to the control group (ie, village B), community health workers, also known as accredited social health activists (ASHAs), visited each pregnant woman and discussed maternal health concerns orally, one-to-one; ensured ambulance availability if needed; and provided financial incentives for women delivering at the hospital. In the intervention group (ie, village A), ASHAs leveraged mHealth (ie, the MfM app) technology to discuss maternal health concerns and measures. All communication occurred in their native language, Hindi. Please note that participants in both groups received the same information, one-to-one; had the same number of visits; and had visits lasting a similar duration.

A total of 800 ASHAs were involved in this study, of which 400 were allocated to the intervention group. ASHAs comprise a community-based health worker group founded as a part of the National Rural Health Mission by the Indian Ministry of Health and Family Welfare. In the intervention group, ASHAs were trained in using the mHealth app (ie, MfM) for 2 days and were equipped with a Nokia phone preinstalled with the MfM app.

ASHAs used the MfM app to engage with pregnant women during each home visit in the intervention group. ASHA staff visited each pregnant woman four times in the prenatal phase (ie, every trimester) and twice during the postnatal phase (ie, on the third and sixth months after childbirth). During each visit, the ASHA workers used the MfM app to counsel pregnant women on antenatal care (ANC), intranatal care, and postnatal care. Figure 1 shows the training of ASHAs and their interaction with pregnant women. The control group also received the same number of AHSA visits; however, ASHA workers did not use the MfM app. Visits in both the intervention and control groups lasted approximately 45 minutes.

Figure 1. Components of the intervention. (A) Training of accredited social health activists. (B) Women using the Mobile for Mothers app as part of the intervention.

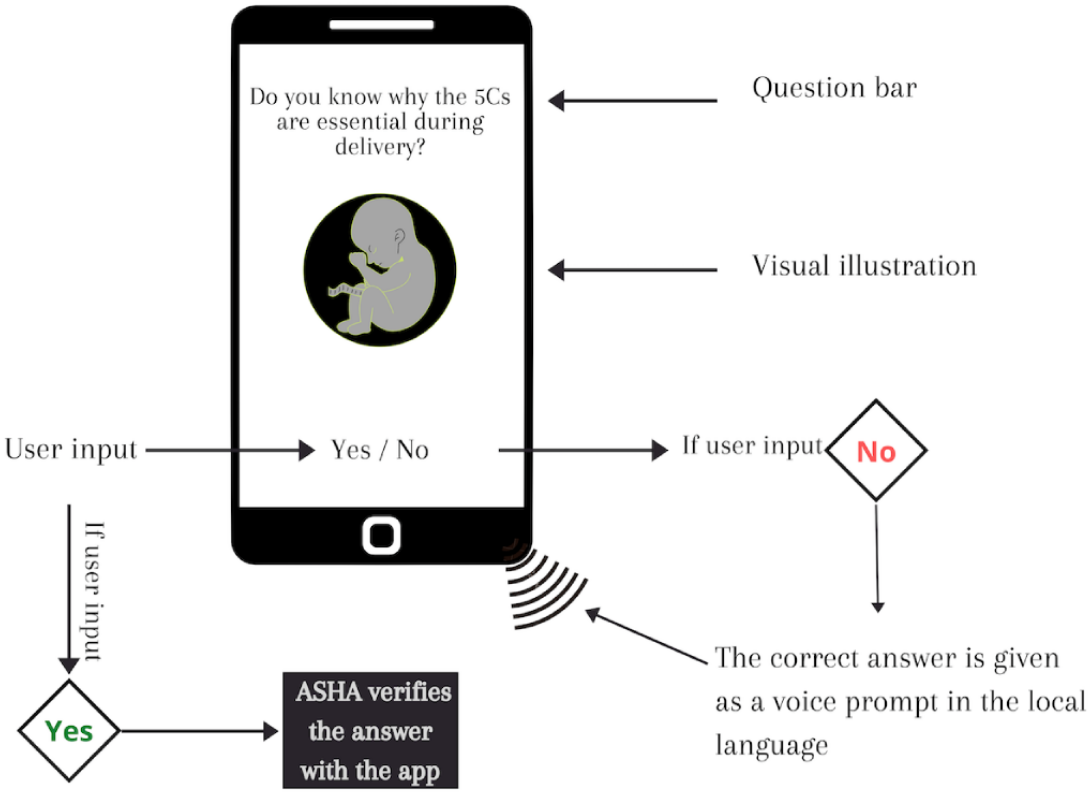
Sample Size

A priori power analysis was completed to estimate the minimum sample size for the study. The a priori power analysis included 2-tailed assumptions, an estimated power of 0.80 [24], an α error probability of .01, and an effect size of 0.2 [25]. The results of the a priori power analysis supported the inclusion of at least 1172 participants in the study.

The Mobile for Mothers App

MfM is a case management solution in rural India for ASHAs. The program helps handle the registration, service, and monitoring of all clients and events related to the ASHAs. MfM collects data from each home visit and transfers the data for service optimization, wellness surveillance, and workflow measures to the NEEDS management committee. MfM is designed for low-literacy users so they can operate the app on affordable Java-enabled phones or Android-based smartphones that run free and open-source apps. It contains registration forms, checklists, tracking of danger signs, and an interactive voice-recording system for instructional prompts. In addition, the app provides maternal health information through texts, photographs, and voice prompts in their native language of Hindi to pregnant women and mothers, as illustrated in Figure 2.

Figure 2. The Mobile for Mothers mobile health app. The original app was in the Hindi language. 5C: five cleans (clean hands, clean place, clean cloths, clean cord cut with a clean blade, and clean cord clamped with a clean thread); ASHA: accredited social health activist.



Data Collection

In January 2014, researchers conducted a complete house list and identified eligible respondents: all pregnant women between 18 and 45 years of age. Soon after, a team of trained project members, who were local to the region, administered a paper-based survey in the control and intervention villages to collect baseline data. During the survey, team members read the questions in Hindi and marked the responses for all the participants, including illiterate and literate pregnant women. The survey used in this study was based on The National Family Health Survey [26] and measured the pregnant women's maternal health and hygiene awareness.

The mHealth intervention commenced soon after the baseline survey was completed in February 2014. The end-line data were collected between November 2015 and January 2016 by the same project members and in the same manner.

Outcome Measures

Maternal health awareness and related behavioral changes were captured in this study. Awareness was calculated as a binary variable where women were deemed to be aware of maternal health information if they responded "yes," followed by the correct explanation, to the following awareness questions:

1. Do you know why the five cleans (5Cs: clean hands, clean place, clean cloths, clean cord cut with a clean blade, and clean cord clamped with a clean thread) are essential during delivery?
2. Pregnant women should receive two tetanus vaccine injections. There should be a difference of 1 month between the first and second injections. Does this protect you and your baby from tetanus?
3. Do you know that if you have problems like painful or burning urination and itchy genitals during pregnancy they are indicative of a reproductive tract infection?
4. Do you know why you need to test for HIV/AIDS during or before pregnancy?
5. Have you consumed all of your iron tablets?

The ASHAs determined the validity of their responses. The MfM app also prompted the correct answers in Hindi.

Data Analysis

Responses from the baseline and end-line paper-based surveys were manually entered into a Microsoft Excel sheet by two research assistants (NEEDS staff). After verifying the data entry, all data were imported into SPSS software (version 27; IBM Corp) for further analysis.

First, we calculated descriptive statistics related to demographics. The percentage changes for each variable between baseline and end line in each group were also calculated. The baseline preintervention groups were compared to the end-line postintervention groups using Pearson chi-square analyses, showing the significance of pre- to postintervention differences (ie, percentage changes) within each group. In addition, Mantel-Haenszel tests for conditional independence were conducted to determine if the pre- to postintervention differences in the intervention group were significantly different (at a 99% CI) from the pre- to postintervention differences in the control group.

Ethics Approval

This study obtained ethical approval from the Institutional Review Board of the Centre for Media Studies, New Delhi, India (approval No. IRB00006230). Given the low level of literacy among the sample population, verbal consent was acquired rather than written documentation. Researchers read the consent form in Hindi. Before the study began, all participants were briefed on the in-depth intent of the study. No participant identifiers were obtained during the study.

Results

The survey consisted of 1480 respondents, with 740 women per group. In total, 73.2% (542/740) and 70.7% (523/740) of the respondents in the intervention and control groups, respectively, belonged to other backward castes. In both groups, about half of the respondents were illiterate. Most women in both study groups were housewives. Table 1 shows the participant demographics.

Table 1. Demographic characteristics of the study participants (N=1480).

Demographic characteristic	Control village (n=740), n (%)	Intervention village, (n=740), n (%)
Caste		
Scheduled caste	131 (17.7)	100 (13.5)
Scheduled tribe	22 (3.0)	29 (3.9)
Other backward castes	523 (70.7)	542 (73.2)
Other than scheduled cast, scheduled tribe, or other backward castes	63 (8.5)	69 (9.3)
Educational level		
Illiterate (never been to school)	384 (51.9)	385 (52.0)
Primary (1-5 years of schooling)	124 (16.8)	141 (19.1)
Secondary (6-10 years of schooling)	199 (26.9)	187 (25.3)
Higher (≥11 years of schooling)	33 (4.5)	29 (3.9)
Occupational status		
Housewife	684 (92.4)	645 (87.2)
Age during intervention (years)		
18-19	135 (18.2)	145 (19.6)
20-24	383 (51.8)	369 (49.9)
25-29	160 (21.6)	158 (21.4)
30-34	47 (6.4)	47 (6.4)
35-45	15 (2.0)	21 (2.8)
Age at marriage (years)		
<18	470 (63.5)	501 (67.7)

As shown in Table 2, except for awareness regarding tetanus vaccine injections, we noted significant improvement, from baseline to end line, in maternal health awareness in both the intervention and control groups. However, the magnitude of improvement was significantly higher in the intervention group.

Awareness regarding the 5Cs in the intervention group increased (P<.001) from among 143 out of 740 participants at baseline to 555 out of 740 participants at end line. In contrast, in the control group, awareness increased (P<.001) from among 108 out of 740 participants at baseline to 555 out of 740 participants at end line. However, the increase in awareness was significantly greater in the intervention group than in the control group (P<.001).

Awareness about tetanus vaccine injections and the fact that pregnant women should receive two shots of tetanus vaccine were significantly increased (P<.001) from among 73 out of 740 participants to 372 out of 555 participants in the intervention group. However, for the control group, awareness increased (P<.001) from among 39 out of 108 participants to 220 out of 492 participants. The magnitude of improvement for tetanus vaccine awareness was also significantly higher in the intervention group than in the control group (P<.001).

In the intervention group, awareness regarding the fact that problems like painful or burning urination and itchy genitals during pregnancy are indicative of a reproductive tract infection (ie, vaginal yeast infection) increased (P<.001) from among 15 out of 740 participants to 608 out of 740 participants. However, in the control group, awareness increased (P=.10) from among 7 out of 740 participants to 132 out of 740 participants. The magnitude of improvement regarding the awareness of reproductive tract infection was also significantly higher in the intervention group than in the control group (P<.001).

Similarly, knowledge about HIV testing increased (P<.001) from among 39 out of 740 participants to 572 out of 740

participants in the intervention group. In contrast, in the control group, knowledge increased from among 28 out of 740 participants to 131 out of 740 participants. The change in HIV-related awareness was significantly greater among the intervention group than among the control group (*P*<.001).

Since awareness can be a precursor to behavioral change, we noted that the number of pregnant women in the intervention group who consumed the prescribed dosage of iron tablets increased (*P*<.001) from 193 out of 288 participants at baseline to 612 out of 663 participants at end line. However, the number of pregnant women in the control group who consumed the prescribed dosage of iron tablets increased (*P*<.001) from 129 out of 212 participants at baseline to 223 out of 297 participants at end line. The magnitude of behavioral improvement for iron consumption was significantly higher in the intervention group than in the control group (*P*<.001).

Table 2. Comparison of responses to awareness questions by the intervention and control group participants.

Question and group	Baseline, n (%)[a]	End line, n (%)[b]	Change from baseline to end line, %	Comparison of baseline and end line		Comparison of change between groups	
				χ^2_1	*P* value	χ^2_1	*P* value
Do you know why the five cleans are essential during delivery?[c]							
Control	108 (14.6)	492 (66.5)	51.9	413.3	<.001	—[d]	—
Intervention	143 (19.3)	555 (75.0)	55.7	460.3	<.001	870.5	<.001
Pregnant women should receive two tetanus vaccine injections. There should be a difference of 1 month between the first and second injections. Does this protect you and your baby from tetanus?[e]							
Control	39 (36.1)	220 (44.7)	8.6	2.7	.10	—	—
Intervention	73 (9.9)	372 (67.0)	57.1	459.4	<.001	380.3	<.001
Do you know that if you have problems like painful or burning urination and itchy genitals during pregnancy they are indicative of a reproductive tract infection?[c]							
Control	7 (0.9)	132 (17.8)	16.9	124.1	<.001	—	—
Intervention	15 (2.0)	608 (82.2)	80.2	974.8	<.001	1055.6	<.001
Do you know why you need to test for HIV/AIDS during or before pregnancy?[c]							
Control	28 (3.8)	131 (17.7)	13.9	74.8	<.001	—	—
Intervention	39 (5.3)	572 (77.3)	72.0	791.9	<.001	804.8	<.001
Have you consumed all of your iron tablets?[f]							
Control	129 (60.8)	223 (75.1)	14.3	11.8	.001	—	—
Intervention	193 (67.0)	612 (92.3)	25.3	98.8	<.001	87.7	<.001

[a]The number of respondents in this column represents the number at baseline who responded "yes" to the questions.

[b]The number of respondents in this column represents the number at end line who responded "yes" to the questions.

[c]Control and intervention group participants at baseline and end line: n=740.

[d]Statistics for group comparisons are reported only in the intervention group row.

[e]Control group participants at baseline: n=108; control group participants at end line: n=492; intervention group participants at baseline: n=740; intervention group participants at end line: n=555.

[f]Control group participants at baseline: n=212; control group participants at end line: n=297; intervention group participants at baseline: n=288; intervention group participants at end line: n=663.

Discussion

Principal Findings

When most people are demonstrating techno-skepticism (ie, a skeptical attitude toward technology), in our study, pregnant women residing in tribal societies exhibited characteristics of the "learned skeptic." However, in augmenting their awareness, these tribes learned maternal health lessons efficiently from the MfM app. In addition, the results of our study indicate the potential of MfM specifically, or mHealth in general, as an educational tool for tribal communities.

Existing literature has acknowledged several mobile apps that can assist with maternal education and support socially disadvantaged pregnant women [27]. Despite having minimal prior experience with mobile devices and no health literacy, our study demonstrates how tribal communities learned information about maternal health and hygiene when delivered through the mHealth app in a user-centered manner (ie, using their regional language and visuals). The MfM intervention also improved the pregnant women's health behavior, in the form of being persistent with iron supplementation. Persistence is measured by how an individual's conduct meets the desired health care objectives jointly identified with the health care professional

[28]. Overall, our study indicates that the mode of information transference can determine its acceptance by users and their likelihood of gaining new knowledge. According to the cognitive theory of multimedia learning, people learn more effectively from words and images than from words alone [29]. This theory can partially explain why tribal communities learned new information through MfM significantly more than through traditional government interventions in which information transference was verbal. Using both auditory and visual channels in MfM (ie, dual-coding theory) perhaps helped pregnant women acquire new knowledge. However, these assumptions require further confirmation.

Maternal Health Awareness

Lack of maternal health awareness is a significant concern in rural and tribal India, causing several health concerns, such as neural tube defects [10] and restricted fetal growth [30]. Poor awareness regarding hygiene (ie, the 5Cs) and tetanus, a life-threatening bacterial disease, have also been responsible for genital tract infection (ie, vaginal yeast infection), puerperal sepsis, and even morbidity among rural and tribal communities [31-33]. Our study demonstrated that mHealth, like MfM, could augment health awareness regarding the 5Cs and tetanus. In our research, more pregnant women in the intervention group acknowledged that the 5Cs can prevent infection in newborn babies and mothers.

Poor awareness has also resulted in tetanus infections and immunodeficiency syndromes, causing harm to pregnant women and newborn children [34,35]. In pregnant women, reproductive tract infection risk predominantly occurs due to a lack of awareness and can lead to postabortal sepsis, puerperal sepsis, and even perinatal death [11]. Deaths of pregnant women due to HIV in rural communities are often induced by poor pregnancy management, primarily due to the lack of knowledge about HIV infection [36].

Our findings were consistent with other studies concerning mHealth and maternal health, where increased awareness was reported. A 2014 study in India's rural region evaluated the impact of mHealth on raising awareness regarding maternal health. The study found that more individuals learned about danger signs during pregnancy, such as infections, after receiving text messages from the mHealth app [37]. Another study in Nigeria investigated the impact of mHealth on maternal knowledge. The study reported that pregnant women without mHealth access had significantly lower knowledge of maternal danger signs than those with access to mHealth apps [38].

Unlike our study, where mHealth increased awareness regarding HIV and tetanus among tribal pregnant women, a randomized controlled trial implementing mHealth to augment maternal health awareness did not find any significant increase in awareness regarding tetanus and HIV [39].

Iron Tablet Consumption

Exposure to an interactive voice recording system for instructional prompts, like MfM, could improve a pregnant woman's health behaviors. For example, in our study, more pregnant women in the intervention group than in the control group consumed the prescribed dosage of iron tablets (25%; $P<.001$). Other studies also reported similar findings. In 2014, a study in Kenya reported that pregnant women (91.6%) receiving an mHealth intervention consumed the necessary dosage of supplements [39]. In 2017, a study in India registered a significant increase in the consumption of iron supplements among pregnant women (81%) receiving mHealth support [40]. Similarly, another study conducted in Indonesia reported a 2.6% increase in the consumption of iron tablets among pregnant women receiving an mHealth intervention [41].

Limitations

This study only focused on two villages; though the sample size was sufficient, our findings may not be generalizable across all tribal and rural communities. Due to limited access to mobile phones in rural India, ASHA workers were responsible for carrying the mobile device with them during each intervention. Pregnant women, being passive users, only used the mobile app in the presence of the ASHAs. Further research is needed to capture the direct impact of mHealth on maternal health awareness and behavior in tribal communities when pregnant women are active users without assistance from trained personnel such as ASHAs.

Implications, Future Directions, and Conclusions

Although improvements in maternal health care awareness were observed in both control and intervention groups, our study provides evidence that mHealth interventions can improve critical maternal health awareness and related behavior among tribal pregnant women at a significantly higher magnitude. Our analysis also exhibited the potential of MfM to minimize the cognitive biases of tribal communities. Tribal communities often show anchoring bias. In other words, their dependency on prior knowledge (ie, "traditional health care system") [42] is significant. Their health care practices and beliefs are primarily determined and constructed by their faith in traditional knowledge, which involves three ways of addressing health complications—natural medicine, psychosomatic treatments, and religious rituals—that typically reject the intervention of scientific methods [42]. However, participants in the intervention group behaved otherwise. Despite firmly believing in the "traditional health care system," tribal communities in the intervention group were noted to break away from their traditional beliefs and adhere to scientific or modern maternal health care practices.

The findings of our study exhibit the effectiveness of mHealth and encourage the use of mHealth as a supporting tool for health workers serving in rural regions. In addition to augmenting maternal health awareness and behavior, mHealth interventions may lower health care costs in several ways, including transportation cost reduction for patients and digitizing data collection (ie, registration of pregnant women, tracking their ANC services, and others). Moreover, in urban societies where pregnant women have access to mobile phones, the internet, and sufficient ability to use smartphones independently, mHealth apps can serve as an alternative source of maternal health information. In 2013, a study reported that about 40% of its 35 participants had used at least one mHealth app for maternal information [43]. Another study in 2016 surveyed 410 women and found that 92% of users found mHealth to be useful for

pregnancy-related information [44]. Therefore, future studies should consider evaluating the economic feasibility of mHealth in maternal health awareness [45].

Acknowledgments

This study was funded by Simavi, Netherlands (grant 3312005), and Deutsche Welthungerhilfe (grant WHHInd/1287). The content is solely the responsibility of the authors.

Data Availability

Deidentified end-line data will be made available only for research purposes upon request.

Authors' Contributions

AC and MC conceptualized the paper. MC provided data, and AC designed the mHealth intervention. AC participated in the analysis and wrote the manuscript draft. MC provided overall guidance. Both authors reviewed the results, facilitated the intervention, and equally contributed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 65 KB - mhealth_v10i9e38368_fig.pdf]

References

1. Jaysawal N. Rural health system in India: A review. Int J Soc Work Hum Serv Pract 2015 Feb;3(1):29-37. [doi: 10.13189/ijrh.2015.030104]
2. Kaur M, Gupta M, Pandara Purayil V, Rana M, Chakrapani V. Contribution of social factors to maternal deaths in urban India: Use of care pathway and delay models. PLoS One 2018;13(10):e0203209 [FREE Full text] [doi: 10.1371/journal.pone.0203209] [Medline: 30300352]
3. Singh A, Padmadas SS, Mishra US, Pallikadavath S, Johnson FA, Matthews Z. Socio-economic inequalities in the use of postnatal care in India. PLoS One 2012;7(5):e37037 [FREE Full text] [doi: 10.1371/journal.pone.0037037] [Medline: 22623976]
4. Singh A, Pallikadavath S, Ram F, Ogollah R. Inequalities in advice provided by public health workers to women during antenatal sessions in rural India. PLoS One 2012;7(9):e44931 [FREE Full text] [doi: 10.1371/journal.pone.0044931] [Medline: 23028688]
5. Horwood G, Opondo C, Choudhury SS, Rani A, Nair M. Risk factors for maternal mortality among 1.9 million women in nine empowered action group states in India: Secondary analysis of Annual Health Survey data. BMJ Open 2020 Aug 20;10(8):e038910 [FREE Full text] [doi: 10.1136/bmjopen-2020-038910] [Medline: 32819952]
6. Gülmezoglu AM, Lawrie TA, Hezelgrave N, Oladapo OT, Souza JP, Gielen M, et al. Interventions to reduce maternal and newborn morbidity and mortality. In: Black R, Laxminarayan R, Temmerman M, Walker N, editors. Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities. 3rd edition (volume 2). Washington, DC: The International Bank for Reconstruction and Development / The World Bank; 2016.
7. Sharma DC. India still struggles with rural doctor shortages. Lancet 2015 Dec 12;386(10011):2381-2382. [doi: 10.1016/S0140-6736(15)01231-3] [Medline: 26700521]
8. Barnes-Josiah D, Myntti C, Augustin A. The “three delays” as a framework for examining maternal mortality in Haiti. Soc Sci Med 1998 Apr;46(8):981-993. [doi: 10.1016/s0277-9536(97)10018-1]
9. Thaddeus S, Maine D. Too far to walk: Maternal mortality in context. Soc Sci Med 1994 Apr;38(8):1091-1110. [doi: 10.1016/0277-9536(94)90226-7] [Medline: 8042057]
10. Narayan L. Addressing language barriers to healthcare in India. Natl Med J India 2013;26(4):236-238. [Medline: 24758452]
11. Taye B, Abeje G, Mekonen A. Factors associated with compliance of prenatal iron folate supplementation among women in Mecha district, Western Amhara: A cross-sectional study. Pan Afr Med J 2015;20:43 [FREE Full text] [doi: 10.11604/pamj.2015.20.43.4894] [Medline: 26090001]
12. Patel DA, Burnett NM, Curtis KM. Reproductive Tract Infections: Reproductive Health Epidemiology Series Module 3. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Reproductive Health; 2003 Jun. URL: https://www.cdc.gov/reproductivehealth/productspubs/pdfs/epi_module_03a_tag508.pdf [accessed 2022-09-13]
13. Mahapatro M, Kalla AK. Health seeking behaviour in a tribal setting. Health Popul Perspect Issues 2000;23(4):160-169.

14. Ugwa E. Nutritional practices and taboos among pregnant women attending antenatal care at general hospital in Kano, Northwest Nigeria. Ann Med Health Sci Res 2016;6(2):109-114 [FREE Full text] [doi: 10.4103/2141-9248.181846] [Medline: 27213094]
15. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
16. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: Where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
17. Becker S, Mitchell A, Königsmann T, Kribben A, Erbel R. [Mobile applications and management of hypertension: possibilities, problems and perspectives] [Article in German]. Herz 2012 Nov;37(7):742-745. [doi: 10.1007/s00059-012-3685-y] [Medline: 23052900]
18. O'Donovan J, Bersin A, O'Donovan C. The effectiveness of mobile health (mHealth) technologies to train healthcare professionals in developing countries: A review of the literature. BMJ Innov 2014 Dec 12;1(1):33-36. [doi: 10.1136/bmjinnov-2014-000013]
19. Mohan D, Bashingwa JJH, Tiffin N, Dhar D, Mulder N, George A, et al. Does having a mobile phone matter? Linking phone access among women to health in India: An exploratory analysis of the National Family Health Survey. PLoS One 2020;15(7):e0236078 [FREE Full text] [doi: 10.1371/journal.pone.0236078] [Medline: 32687527]
20. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
21. Choudhury A, Asan O, Choudhury MM. Mobile health technology to improve maternal health awareness in tribal populations: Mobile for Mothers. J Am Med Inform Assoc 2021 Oct 12;28(11):2467-2474 [FREE Full text] [doi: 10.1093/jamia/ocab172] [Medline: 34459478]
22. Ilozumba O, Dieleman M, Kraamwinkel N, Van Belle S, Chaudoury M, Broerse JEW. "I am not telling. The mobile is telling": Factors influencing the outcomes of a community health worker mHealth intervention in India. PLoS One 2018;13(3):e0194927 [FREE Full text] [doi: 10.1371/journal.pone.0194927] [Medline: 29584773]
23. Ilozumba O, Van Belle S, Dieleman M, Liem L, Choudhury M, Broerse JEW. The effect of a community health worker utilized mobile health application on maternal health knowledge and behavior: A quasi-experimental study. Front Public Health 2018;6:133 [FREE Full text] [doi: 10.3389/fpubh.2018.00133] [Medline: 29868541]
24. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
25. Brydges CR. Effect size guidelines, sample size calculations, and statistical power in gerontology. Innov Aging 2019 Aug;3(4):igz036 [FREE Full text] [doi: 10.1093/geroni/igz036] [Medline: 31528719]
26. National Family Health Survey - 4 2015-16: State Fact Sheet—Jharkhand. New Delhi, India: Ministry of Health and Family Welfare, Government of India URL: http://rchiips.org/nfhs/pdf/NFHS4/JH_FactSheet.pdf [accessed 2021-01-21]
27. Dalton JA, Rodger D, Wilmore M, Humphreys S, Skuse A, Roberts CT, et al. The Health-e Babies App for antenatal education: Feasibility for socially disadvantaged women. PLoS One 2018;13(5):e0194337 [FREE Full text] [doi: 10.1371/journal.pone.0194337] [Medline: 29768407]
28. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: Terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]
29. Mayer RE. Cognitive theory of multimedia learning. In: The Cambridge Handbook of Multimedia Learning. Cambridge, UK: Cambridge University Press; 2005:31-48.
30. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, Maternal and Child Nutrition Study Group. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet 2013 Aug 03;382(9890):427-451. [doi: 10.1016/S0140-6736(13)60937-X] [Medline: 23746772]
31. Clark J. The global push for institutional childbirths--In unhygienic facilities. BMJ 2016 Mar 10;352:i1473. [doi: 10.1136/bmj.i1473] [Medline: 26966184]
32. Verma R, Khanna P, Dhankar M. Vaccination during pregnancy: Today's need in India. Hum Vaccin Immunother 2016 Mar 03;12(3):668-670 [FREE Full text] [doi: 10.1080/21645515.2015.1093265] [Medline: 26619155]
33. Yaya S, Kota K, Buh A, Bishwajit G. Prevalence and predictors of taking tetanus toxoid vaccine in pregnancy: A cross-sectional study of 8,722 women in Sierra Leone. BMC Public Health 2020 Jun 05;20(1):855 [FREE Full text] [doi: 10.1186/s12889-020-08985-y] [Medline: 32503478]
34. Mayekiso ND, Ter Goon D, Vellem N, Okafor UB, Rala NM. Perceptions of women of reproductive age towards maternal death in Qaukeni sub-District, Eastern Cape Province, South Africa: A qualitative study. Afr J Reprod Health 2020 Dec;24(4):147-163. [doi: 10.29063/ajrh2020/v24i4.16] [Medline: 34077080]
35. Hill WC, Lindsay MK, Green VL. Executive summary and call to action: Maternal Mortality Summit at National Medical Association Meeting, July 28, 2019 in Honolulu, Hawaii Obstetrics and Gynecology Section of the National Medical Association. J Natl Med Assoc 2020 Aug;112(4):402-410. [doi: 10.1016/j.jnma.2020.04.013] [Medline: 32561137]

36. Lathrop E, Jamieson DJ, Danel I. HIV and maternal mortality. Int J Gynaecol Obstet 2014 Nov;127(2):213-215 [FREE Full text] [doi: 10.1016/j.ijgo.2014.05.024] [Medline: 25097142]
37. Datta SS, Ranganathan P, Sivakumar KS. A study to assess the feasibility of text messaging service in delivering maternal and child healthcare messages in a rural area of Tamil Nadu, India. Australas Med J 2014;7(4):175-180 [FREE Full text] [doi: 10.4066/AMJ.2014.1916] [Medline: 24817911]
38. Jennings L, Omoni A, Akerele A, Ibrahim Y, Ekanem E. Disparities in mobile phone access and maternal health service utilization in Nigeria: A population-based survey. Int J Med Inform 2015 May;84(5):341-348. [doi: 10.1016/j.ijmedinf.2015.01.016] [Medline: 25737460]
39. Fedha T. Impact of mobile telephone on maternal health service care: A case of Njoro Division. Open J Prev Med 2014;04(05):365-376. [doi: 10.4236/ojpm.2014.45044]
40. Bangal VB, Borawake SK, Gavhane SP, Aher KH. Use of mobile phone for improvement in maternal health: A randomized control trial. Int J Reprod Contracept Obstet Gynecol 2017 Nov 23;6(12):5458. [doi: 10.18203/2320-1770.ijrcog20175260]
41. Anitasari D, Andrajati R. Effectiveness of short message service reminders and leaflets in complying with iron supplementation in pregnant women in Depok City, Indonesia. Asian J Pharm Clin Res 2017 Oct 01;10(17):42. [doi: 10.22159/ajpcr.2017.v10s5.23092]
42. Balgir R. Tribal health problems, disease burden and ameliorative challenges in tribal communities with special emphasis on tribes of Orissa. In: Proceedings of the National Symposium on Tribal Health. 2006 Presented at: National Symposium on Tribal Health; October 19-20, 2006; Jabalpur, India p. 161-176 URL: https://www.nirth.res.in/publications/nsth/22.RS.Balgir.pdf
43. Rodger D, Skuse A, Wilmore M, Humphreys S, Dalton J, Flabouris M, et al. Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia. Aust J Prim Health 2013;19(4):308. [doi: 10.1071/py13029]
44. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
45. Bergmo TS. How to measure costs and benefits of eHealth interventions: An overview of methods and frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]

Abbreviations

5Cs: five cleans (clean hands, clean place, clean cloths, clean cord cut with a clean blade, and clean cord clamped with a clean thread)
ANC: antenatal care
ASHA: accredited social health activist
MfM: Mobile for Mothers
mHealth: mobile health
NEEDS: Network for Enterprise Enhancement and Development Support
NGO: nongovernmental organization

Edited by L Buis; submitted 29.03.22; peer-reviewed by C Memering, YM Huang; comments to author 10.06.22; revised version received 02.07.22; accepted 02.09.22; published 21.09.22.

Please cite as:
Choudhury A, Choudhury M
Mobile for Mothers mHealth Intervention to Augment Maternal Health Awareness and Behavior of Pregnant Women in Tribal Societies: Randomized Quasi-Controlled Study
JMIR Mhealth Uhealth 2022;10(9):e38368
URL: https://mhealth.jmir.org/2022/9/e38368
doi:10.2196/38368
PMID:36129749

Original Paper

Evaluation of Diagnostic and Triage Accuracy and Usability of a Symptom Checker in an Emergency Department: Observational Study

Hamish S F Fraser[1,2], BSc, MBChB, MS; Gregory Cohan[3], BA; Christopher Koehler[4], BA, MMS, MD; Jared Anderson[4], BA, MD; Alexis Lawrence[5], BA, MD; John Pateña[6], MA, MPH; Ian Bacher[1], BA, PhD; Megan L Ranney[2,4,6], BA, MPH, MD

[1]Brown Center for Biomedical Informatics, Warren Alpert Medical School, Brown University, Providence, RI, United States

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

[3]Warren Alpert Medical School, Brown University, Providence, RI, United States

[4]Department of Emergency Medicine, Brown University, Providence, RI, United States

[5]Harvard Medical Faculty Physicians, Department of Emergency Medicine, St Luke's Hospital, New Bedford, MA, United States

[6]Brown-Lifespan Center for Digital Health, Providence, RI, United States

Corresponding Author:
Hamish S F Fraser, BSc, MBChB, MS
Brown Center for Biomedical Informatics
Warren Alpert Medical School
Brown University
233 Richmond Street
Providence, RI, 02912
United States
Phone: 1 401 863 1815
Email: hamish_fraser@brown.edu

Abstract

Background: Symptom checkers are clinical decision support apps for patients, used by tens of millions of people annually. They are designed to provide diagnostic and triage advice and assist users in seeking the appropriate level of care. Little evidence is available regarding their diagnostic and triage accuracy with direct use by patients for urgent conditions.

Objective: The aim of this study is to determine the diagnostic and triage accuracy and usability of a symptom checker in use by patients presenting to an emergency department (ED).

Methods: We recruited a convenience sample of English-speaking patients presenting for care in an urban ED. Each consenting patient used a leading symptom checker from Ada Health before the ED evaluation. Diagnostic accuracy was evaluated by comparing the symptom checker's diagnoses and those of 3 independent emergency physicians viewing the patient-entered symptom data, with the final diagnoses from the ED evaluation. The Ada diagnoses and triage were also critiqued by the independent physicians. The patients completed a usability survey based on the Technology Acceptance Model.

Results: A total of 40 (80%) of the 50 participants approached completed the symptom checker assessment and usability survey. Their mean age was 39.3 (SD 15.9; range 18-76) years, and they were 65% (26/40) female, 68% (27/40) White, 48% (19/40) Hispanic or Latino, and 13% (5/40) Black or African American. Some cases had missing data or a lack of a clear ED diagnosis; 75% (30/40) were included in the analysis of diagnosis, and 93% (37/40) for triage. The sensitivity for at least one of the final ED diagnoses by Ada (based on its top 5 diagnoses) was 70% (95% CI 54%-86%), close to the mean sensitivity for the 3 physicians (on their top 3 diagnoses) of 68.9%. The physicians rated the Ada triage decisions as 62% (23/37) *fully agree* and 24% (9/37) *safe but too cautious*. It was rated as *unsafe and too risky* in 22% (8/37) of cases by at least one physician, in 14% (5/37) of cases by at least two physicians, and in 5% (2/37) of cases by all 3 physicians. Usability was rated highly; participants *agreed* or *strongly agreed* with the 7 Technology Acceptance Model usability questions with a mean score of 84.6%, although "satisfaction" and "enjoyment" were rated low.

Conclusions: This study provides preliminary evidence that a symptom checker can provide acceptable usability and diagnostic accuracy for patients with various urgent conditions. A total of 14% (5/37) of symptom checker triage recommendations were

deemed unsafe and too risky by at least two physicians based on the symptoms recorded, similar to the results of studies on telephone and nurse triage. Larger studies are needed of diagnosis and triage performance with direct patient use in different clinical environments.

(JMIR Mhealth Uhealth 2022;10(9):e38364) doi:10.2196/38364

KEYWORDS

mobile health; mHealth; symptom checker; diagnosis; user experience

Introduction

Background

Improving medical diagnosis is a high priority, with evidence that the average American will experience at least one important misdiagnosis in their lifetime and that 5% of outpatient diagnoses are incorrect [1]. Although a number of initiatives have sought to improve outpatient diagnosis by physicians and other health care workers [2,3], less attention has been paid to the key role patients play in ensuring they receive effective and timely diagnosis and treatment [4]. Late recognition of many diseases can lead to poor outcomes, whether for acute diagnoses such as myocardial infarction and stroke or for more chronic diseases, including renal failure [5] and carcinomas [6]. Diagnostic and triage apps designed for patients, often called symptom checkers, have become widely available to the general public over the last decade [7]. Leading symptom checkers claim millions of users annually; for example, iTriage claimed 50 million users per year in 2015 [8], and WebMD claimed 4 million users per month in 2019 [9].

These apps typically require the user to enter limited demographic information followed by their chief complaint or symptom. They then ask follow-up questions on the symptoms, which vary in number and strategy by app [10,11]. The output is one or more diagnoses or a triage level and may include suggestions for actions that the user should take, including seeking routine or urgent care. Symptom checker apps differ from diagnostic tools for health care workers; in most cases, symptom checkers do not use data on vital signs, physical examination, current medications, or investigations [12,13].

The context of a person using a symptom checker app in a home or community setting can be similar to calling an urgent care helpline such as the NHS111 service in the United Kingdom [14] (although likely involving less urgent conditions). However, it does not include access to the patient's care record or to human assistance in navigating the algorithm. App use is typically promoted by private companies that develop them; academic developers of apps [15]; health systems that have developed their own symptom checkers (eg, Mayo Clinic [16]); or health systems that have partnered with companies, such as Babylon with the National Health Service in the United Kingdom [17] and Ada with Sutter Health in the United States [18]. Symptom checkers have the potential to help patients identify the correct diagnosis for their problem and the appropriate action to take in seeking care. Symptom checkers could be particularly helpful in improving care for patients with limited access to health systems, such as in rural areas and other underserved communities worldwide. These apps could also assist people uncertain of the significance of symptoms with potentially serious underlying causes, such as chest pain or headache. Alternatively, a symptom checker might miss important diagnoses, discourage users from seeking urgent care, or overwhelm health systems with patients who have nonurgent problems. The latter issue was observed with the phone triage system NHS111 [19] and in a study of the use of telehealth consultations that increased patient contacts rather than just displacing in-person care [20]. Symptom checkers have seen extensive use during the COVID-19 pandemic, with evaluation studies showing good diagnostic accuracy for some systems but significant underdiagnosis in some nationally deployed COVID-19 symptom checker apps [21]. However, many patients with COVID-19 have few or no symptoms, limiting potential sensitivity in the absence of additional data such as pulse oximetry.

Despite ample business promotion of symptom checkers, little rigorous evidence supports their effectiveness, safety, accuracy, ability to decrease the load on health systems, or usability by the full range of users or patients [7,22]. Most studies to date have used clinical vignettes—patient histories with "correct" diagnoses created by physicians—to evaluate key metrics [12,13,23,24]. Although such studies have played an important role in identifying gaps in coverage or weaknesses in diagnostic algorithms, they do not reproduce the experience of patients using a symptom checker. In addition, the vignettes may be less challenging to diagnose than real patient histories collected in an emergency situation, as illustrated in a recent study that included vignettes created using actual patient presentations to an urgent health care hotline [13]. To date, studies that have been based on direct patient use of symptom checkers have enrolled few acute patients or had poorly defined study populations and outcomes [25,26]. As symptom checkers are designed to be used in a home or community setting without direct health care support, to address the critical question of patient safety, it is necessary to enroll patients with serious and potentially life-threatening diseases. For example, in a study of routine symptom checker use in a health system in California, 29% of assessments were for patients considered by the clinical team to be high urgency [18].

Ada Health Symptom Checker

The Ada Health symptom checker's diagnosis algorithm was developed with the original goal of assisting clinicians with the diagnosis of rare diseases. Since the launch of the symptom checker in 2016, its use has grown rapidly in Germany, the United Kingdom, and the United States. A total of 11 million users have carried out 23 million health assessments. It is available in 11 languages and, in 2020, it was rated as the most commonly used symptom checker in 150 countries [27]. On first use of the Ada app, the user is questioned about their age,

sex, gender, and a limited number of pre-existing diseases. They are then asked for their chief complaint. A series of questions then follows in the manner of a "chatbot." Upon completion of the question-and-answer phase, the user is given a list of 1 to 5 "condition suggestions" equivalent to a differential diagnosis and a recommendation for the level of urgency with which to seek care. The underlying algorithm is a Bayesian network. Previous studies of Ada have shown good levels of performance on a wide range of diagnoses using clinical vignettes, including a large study of 8 symptom checkers using 200 vignettes (led by a team from Ada Health with support in design and analysis from outside experts, including HF) [13] and a study by Ceney et al [10], which was independent from Ada. The choice of Ada for this study was based on its widespread use in many countries, preliminary evidence of strong performance, and a willingness to collaborate with outside, independent research teams, as shown by the wide range of published evaluation studies [13,28,29]. The authors also considered other symptom checkers, including conversations with YourMD Ltd, and have independently tested the Ada, Isabel, and WebMD symptom checkers on the diagnosis of chest pain [30].

The study was designed to recruit patients seeking urgent care in an emergency department (ED).

The research questions for this study were as follows: (1) Could patients presenting with an urgent clinical problem effectively record their symptoms and did they find Ada easy to use? (2) Did patient characteristics affect their successful use of the app? (3) What was the sensitivity of Ada for the diagnoses of the ED physicians who saw the patients? (4) Were the diagnoses suggested by Ada as sensitive to the ED physicians' diagnoses as the diagnoses suggested by physicians using the same clinical data? (5) Were the diagnoses suggested by Ada considered reasonable by the physicians? (6) Were the triage suggestions from Ada considered reasonable and safe by the physicians? (7) Did access to vital sign data improve the diagnostic performance of the reviewing physicians?

Methods

Overview

The symptom checker was tested with direct use by patients coming to the Rhode Island Hospital (RIH) ED with a wide range of presenting complaints. The sample population was designed to include "patients with acute or serious medical conditions presenting to an emergency department." Patients were eligible if they were English-speaking, aged ≥18 years, presenting for emergency evaluation of a medical (nontrauma and non–mental health) problem, and deemed by the triage nurse to not be critically ill (Emergency Severity Index score of 2-5). In addition, they had to be able to consent and complete the symptom checker assessment before physician evaluation.

Participants were approached after initial nurse triage but before physician assessment by a research assistant (CK) on a convenience sample of shifts. After obtaining consent from the participants, they were provided with an iPad with the study software installed and followed the symptom checker prompts. Upon completion of the symptom checker questions, they were then asked to complete a usability survey using REDCap (Research Electronic Data Capture; Vanderbilt University), a Health Insurance Portability and Accountability Act–compliant survey software [31]. The survey questions were of three types: (1) demographic data, including age, sex, race, ethnicity, and socioeconomic status; (2) six questions on the participants' use of web-based resources and health information–seeking behavior; and (3) eleven questions from the Technology Acceptance Model [32]. Participants were compensated with US $20 for their time. The version of Ada used in the study used the same diagnostic algorithm as the production system but returned the results as a PDF file emailed to the study team; the patients did not see the results.

Symptom checker data were compiled into deidentified files. Each file included the patient's answers to the symptom checker questions, up to five diagnoses from the symptom checker (termed "condition suggestions" by Ada), and a recommended triage action for the patient (termed "urgency advice" by the Ada app). An example is shown in Multimedia Appendix 1. No symptom checker data were seen by the patient or the physician caring for them. Subsequently, the discharge diagnoses, vital signs, physical examinations, and laboratory or other study results from the ED visit were abstracted from the electronic health record (EHR). If the ED discharge diagnoses were unclear, an attending ED physician independent of the patient's care adjudicated.

To assess performance, 3 independent, board-certified emergency physicians reviewed the data summaries generated by the symptom checker, blinded to the ED diagnoses. REDCap was used to present the data to allow each physician to complete the following tasks in sequence: (1) review the patient-reported symptom checker symptom data and generate their own differential diagnosis and triage level without access to the Ada diagnosis or condition suggestion or the patient's chart, (2) review the patient's vital signs and then restate their top 3 diagnoses and triage levels, and (3) review and critique the Ada diagnoses and triage levels for the case. The results from the 3 ED physicians' critiques of Ada were combined to create an assessment of the appropriateness of the diagnoses and triage levels by majority voting.

The 95% CIs were calculated using the proportions method, and the diagnostic accuracy results were compared using the chi-square test. Interrater agreement between the symptom checker, discharge diagnoses, and independent physician diagnoses was calculated by comparing their percentage of agreement on the ED diagnoses (ie, in what percentage of cases did 2 clinicians match the same ED diagnosis). Clinicians were then compared with Ada in the same fashion, limiting Ada to its top 3 diagnoses. The *comprehensiveness* and *relevance* metrics were calculated for Ada and the physicians to account for multiple diagnoses in each list. Comprehensiveness was calculated as the percentage of ED diagnoses matched by the differential diagnoses of Ada or the physicians (similar to sensitivity). Relevance was calculated as the percentage of the diagnostician's (Ada or the physicians) diagnoses that matched the ED diagnoses (similar to the positive predictive value) [33]. Free-text comments were analyzed thematically by 2 authors (HSFF and GC), and differences were resolved through

discussion. For the user survey, descriptive statistics were calculated for demographic and Likert-scale data using Microsoft Excel.

Ethics Approval

The study was approved by the Interventional Review Board, Research Data Protection Office, Lifespan Healthcare, Providence, Rhode Island (1439681-3). Institutional review board approval was obtained before initiation of the study.

Results

Overview

Over 5 days in September 2019 to October 2019, a total of 143 patients presented to the ED and were screened, and 84 (58.7%) were potentially eligible. Of these 84 patients, 50 (60%) were approached, and 40 (48%) consented. Figure 1 shows the reasons for exclusion. Of the consenting participants, 65% (26/40) were women, with a mean age of 39.3 (SD 15.9; range 18-76) years. Table 1 shows the breakdown by education level and receipt of public assistance for the study and by race and ethnicity for the study and for all patients seen in the ED between September 2019 and October 2019. The study population had a higher proportion of female patients, had a younger mean age, and was more diverse, with a higher proportion who identified as Hispanic and more patients from less common racial and ethnic groups.

Of the 40 patients enrolled, 7 (18%) had incomplete data: 2 (29%) were missing an ED assessment as the patients left against advice, symptom checker assessment files were not generated by Ada in 3 (43%) cases, and Ada assessment files had symptom data and triage information but no diagnosis (with one of these also missing an ED assessment) in 3 (43%) cases. These problems were related to the research environment for Ada, not the symptom checker itself, and were resolved for a subsequent study in primary care. Therefore, overall, 83% (33/40) of the cases had both a full Ada assessment and an ED assessment.

Of these 33 complete cases, 22 (67%) had a clear discharge diagnosis, and the other 11 (33%) had a symptom listed as the final diagnosis. Of these 11 cases, 6 (55%) had a diagnosis of "chest pain," and 2 (18%) had a diagnosis of "back pain." The other 3 cases had poorly specified symptoms: 2 (67%) of "abdominal pain" and 1 (33%) of "dizziness." Consequently, diagnostic accuracy was measurable based on the ED assessment for 30 cases (clear ED diagnosis, myocardial infarction screen, or back pain). There was a mean of 2.5 diagnoses per case (range 1-6) based on the ED record. The review of diagnoses by the 3 independent physicians included 33 cases with complete data, and their review of triage accuracy included all 37 cases with triage data from Ada. The patients were seen in a major ED and level-1 trauma center. Considering the 33 cases, all 6 (18%) patients with chest pain were screened for acute myocardial infarction (AMI). Of the 6 patients, 1 (17%) was admitted with cardiac ischemia, 1 (3%) had a head injury and concussion—possible intercranial hemorrhage—and 1 (3%) had acute appendicitis. The details of the 40 cases, including presenting complaint, ED diagnoses, disposition, and if there was missing data, are provided in Multimedia Appendix 2. One of the reviewing physicians also acted as an expert opinion on uncertain ED diagnoses >6 months after reviewing the Ada data. Figure 2 shows the primary evaluation of the diagnoses from Ada and the physicians with the ED diagnosis.

Figure 1. Patient recruitment and reasons for exclusion. Owing to technical problems, 37 cases were usable for analysis of triage, and 33 cases were usable for analysis of diagnoses.

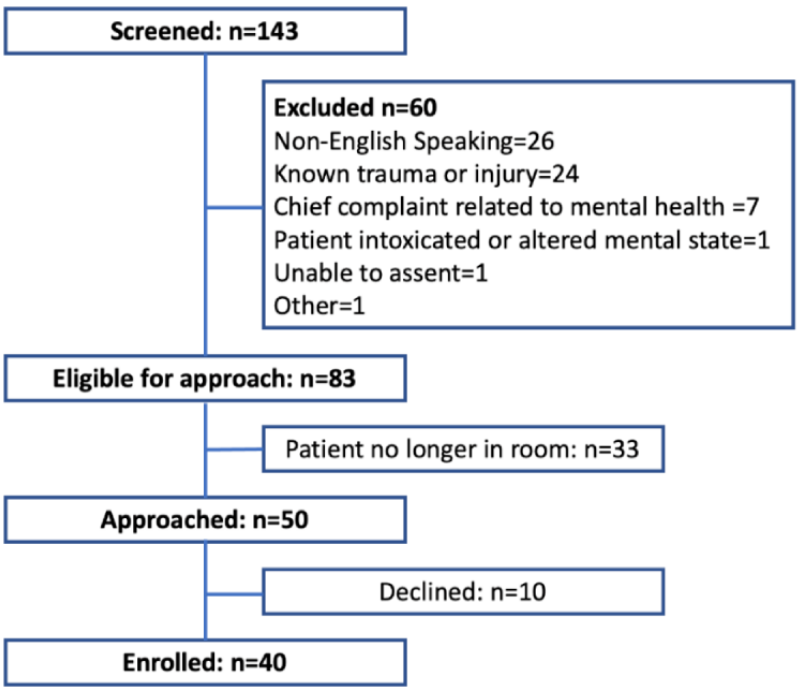


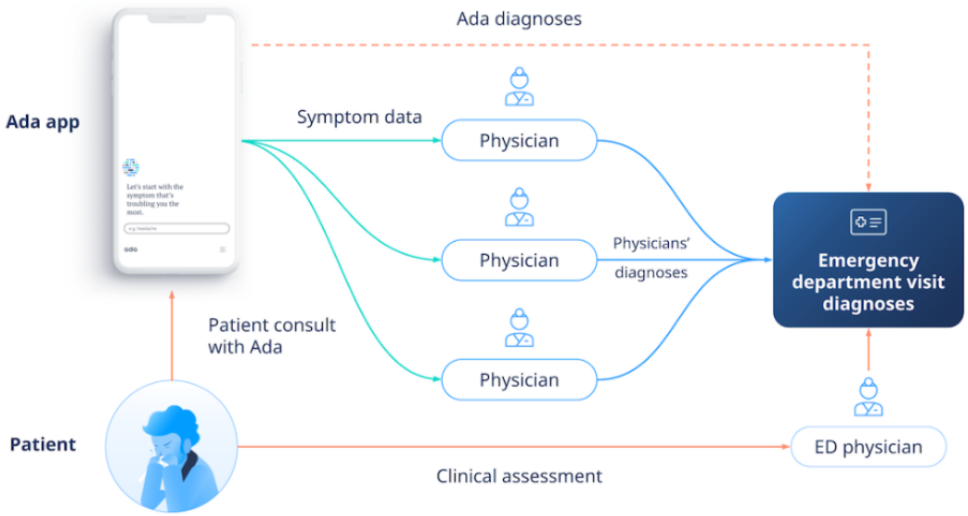
Table 1. Breakdown by self-reported race, education level, and receipt of public assistance (some identified as 2 races). Data from the 16,708 general emergency department (ED) patients seen between September 2019 and October 2019 are shown for comparison. The category "other" in the ED data mapped 88% to "Hispanic or Latino" (N=40).

Characteristic	This study, n (%)	General ED (total patients=16,708), n (%)
Race		
American Indian or Alaska Native	2 (5)	37 (0.22)
Asian	3 (8)	238 (1.42)
Native Hawaiian or other Pacific Islander	1 (3)	36 (0.22)
Black or African American	5 (13)	2203 (13.19)
White	27 (68)	9906 (59.29)
Other	4 (10)	3886 (23.26)
Prefer not to answer	1 (3)	51 (0.31)
Ethnicity—Identified as Hispanic or Latino	19 (48)	4084 (24.44)
Education level		
Some high school	3 (8)	—[a]
High school degree or equivalent (eg, GED[b])	8 (20)	—
Some college	10 (25)	—
Trade or technical or vocational training	3 (8)	—
Associate's degree	5 (13)	—
Bachelor's degree	7 (18)	—
Master's degree	4 (10)	—
Receipt of public assistance		
Yes	10 (25)	—
No	28 (70)	—
Preferred not to say	2 (5)	—

[a]Education level and receipt of public assistance was not recorded in general ED population.

[b]GED: General Educational Development.

Figure 2. Primary comparison of the diagnoses from Ada and the physicians reviewing the Ada data with the emergency department (ED) physician diagnosis.



Metric 1: Comparing Ada Diagnoses With ED Diagnoses

Diagnostic accuracy was measured by the number of Ada diagnoses that matched at least one of the ED discharge diagnoses. For the sensitivity analysis, we considered whether one of the ED discharge diagnoses was Ada's top 1 diagnosis, whether it was in the top 3, or whether it was in the top 5. The top 1 diagnosis matched in 30% (9/30; 95% CI 14%-46%) of cases, the top 3 diagnoses matched in 63% (19/30; 95% CI 46%-81%) of cases, and the top 5 diagnoses matched in 70% (21/30; 95% CI 54%-86%) of cases.

Metric 2: Comparing Independent Physician Diagnoses With ED Diagnoses

The independent physicians were asked to provide up to 3 diagnoses after reviewing symptom checker data with and without vital signs but were blinded to Ada's results. The mean percentage match between diagnoses for the 3 physicians was the top 1 match in 47% (14/30; 95% CI 36%-57%) of cases and the top 3 matches in 69% (20.7/30; 95% CI 59%-78%) of cases. For *physician 1*, the top diagnosis matched the ED diagnosis in 40% (12/30; 95% CI 23%-58%) of cases, and the top 3 diagnoses matched in 70% (21/30; 95% CI 54%-86%) of cases. For *physician 2*, the top diagnosis matched in 57% (17/30; 95% CI 39%-74%) of cases, and the top 3 diagnoses matched in 70% (21/30; 95% CI 54%-86%) of cases. For *physician 3*, the top diagnosis matched in 43% (13/30; 95% CI 26%-61%) of cases, and the top 3 diagnoses matched in 67% (20/30; 95% CI 50%-84%) of cases.

The comparison of the top 1 match for Ada (metric 1) with the combined top 1 matches for the 3 physicians (metric 2) was not significant (P=.07). The results of the top 1 matching diagnosis for *physician 2* showed a significantly higher performance than Ada (P=.02). Matching performance on the top 3 diagnoses was not significantly different between Ada and any of the physicians (Figure 3).

Table 2 shows the percentage of agreement among the pairs of clinicians and clinicians paired with Ada, showing the percentage of cases in which they matched the same ED diagnosis on their top 3 diagnoses. Overall, there was higher agreement between pairs of physicians than between physicians and Ada, but these differences were not statistically significant. The mean level of agreement for Ada was 57% if the top 5 diagnoses were included.

Figure 3. Percentage of cases with at least one match to the final emergency department (ED) diagnosis (MD=physician).

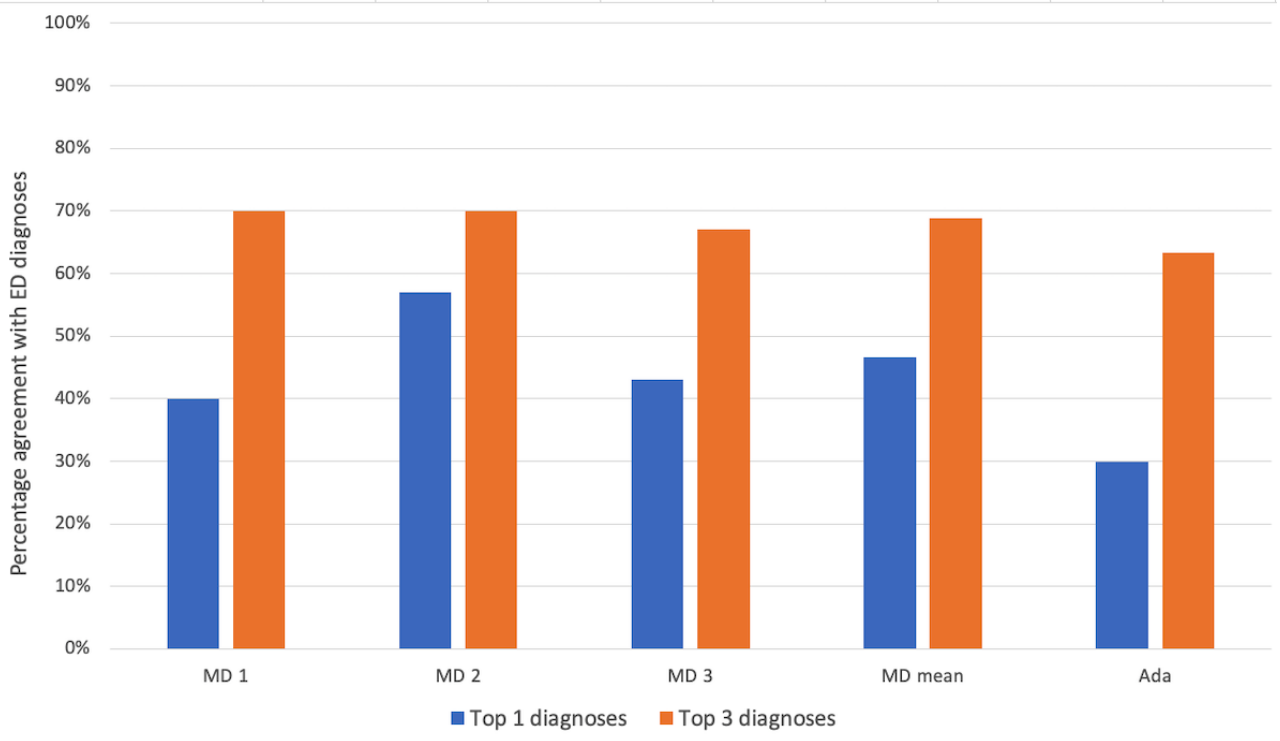


Table 2. Pairwise comparisons of percentage of agreement between physicians and with symptom checker diagnoses to assess interrater agreement (N=30).

Pair number	Pair	Agreement, n (%)
1	Physician 1-physician 2	19 (63)
2	Physician 1-physician 3	17 (57)
3	Physician 1-Ada	13 (43)
4	Physician 2-physician 3	17 (57)
5	Physician 2-Ada	17 (57)
6	Physician 3-Ada	15 (50)
7	Physician 1-ED[b] diagnoses	20 (67)
8	Physician 2-ED diagnoses	21 (70)
9	Physician 3-ED diagnoses	22 (73)
10	Ada-ED diagnoses	19 (63)

[a]Ada mean (rows 3, 5, and 6): 15 (50%); physician mean (rows 1, 2, and 4): 18 (60%); overall mean: 53%.

[b]ED: emergency department.

Metric 3: Evaluating Reasonableness of Ada Diagnoses and Triage Urgency Through Independent Physicians

There were a total of 130 Ada diagnoses; these were reviewed by the 3 clinicians. On the basis of agreement of at least two reviewers, the results were 39% (50/130) "very reasonable," 30% (39/130) "reasonable," 16.2% (21/130) "neither reasonable nor unreasonable," 15% (19/130) "unreasonable," and <1% (1/130) "very unreasonable." The diagnosis considered very unreasonable was age-related farsightedness in a patient presenting with chest pain and headache. Other Ada diagnoses for this case were eye strain, hypertensive emergency, high blood pressure, and low blood sugar. See Table 3 for other examples and additional information on the symptoms entered and associated diagnoses.

Table 3. Diagnoses considered *very unreasonable* by a physician reviewer and *unreasonable* by one or more. The italicized diagnosis was rated as *very unreasonable* by 2 reviewers (additional symptoms were reported to Ada).

Conditions critiqued in Ada diagnoses	Chief complaint (ED[a])	Additional information
Heatstroke	Dizziness, nausea, and headache	Other Ada diagnoses: low blood sugar and viral stomach bug, considered reasonable
Age-related farsightedness	Chest pain and headache	Symptoms entered into Ada included eye strain, reduced vision, laterality (both eyes), and no sudden onset
Hereditary angioedema	Left upper quadrant abdominal pain	Top Ada diagnosis: pancreatitis, considered reasonable or very reasonable
Hereditary angioedema, abdominal wall hematoma, or Meckel diverticulum	Abdominal pain	Top Ada diagnosis: appendicitis, matched by all 3 physicians

[a]ED: emergency department.

When the clinicians assessed Ada's suggested triages, 62% (23/37) were rated as *fully agree*, 24% (9/37) were rated as *safe but too cautious*, and 14% (5/37) were rated as *unsafe and too risky*. A total of 22% (8/37) were rated as *unsafe and too risky* by at least one physician, but only 5% (2/37) were found so by 100% (3/3) of the physicians. Each case report generated by Ada had an overall urgency level; this was normally based on the diagnosis with the most urgent level of triage. However, 11% (4/37) cases had a more urgent triage level for diagnoses ranked at lower probabilities. Including those more urgent triage levels in the analysis reduces the rate of undertriage supported by at least two physicians to 5% (2/37) but does not change overtriage, increasing the *fully agree* category to 70% (26/37). The details of the reviews of triage and diagnoses in the *unsafe and too risky* category are shown in Multimedia Appendix 3.

Metric 4: Comprehensiveness and Relevance Results

Compared with the ED diagnoses, the mean *comprehensiveness* for Ada's top 3 diagnoses was 41%, and the mean *relevance* was 22%. For the top 5 diagnoses, they were 46% and 24%, respectively. Considering only the top 1 diagnosis, Ada's *relevance* was 33%. The mean *comprehensiveness* for the physicians was 46% (range 41%-54%), and the mean *relevance* was 27.7% (range 25%-32%). The mean *relevance* for the physicians' top 1 diagnosis was 47.8% (range 40%-60%).

User Survey

All 40 participants successfully completed the user survey (and were not given access to the Ada results). The results of the survey are shown in Tables 4 and 5. Overall, they reported a high level of use of cell phones to send SMS text messages and enter and view data; however, only 53% (21/40) used a computer

at work. When seeking medical advice, the most frequent source was doctors (physicians; 22/40, 55%), with web-based sources being second at 40% (16/40). A total of 83% (33/40) searched for medical symptoms on the web at least sometimes, and 28% (11/40) had previously used a symptom checker. Regarding overall satisfaction, participants were evenly split among *satisfied*, *neutral*, and *unsatisfied*. Only 23% (9/40) said that the use of Ada was "enjoyable"; however, 70% (28/40) were *likely* or *very likely* to recommend it to a friend.

Participants *agreed* or *strongly agreed* with the following statements (based on the Technology Acceptance Model): *Using Ada would enable me to record my medical symptoms and problems quickly* (33/40, 83%), *Learning to use Ada would be easy for me* (37/40, 93%), *I would find it easy to get Ada to do what I want it to do* (26/40, 65%), *The way to use Ada was clear and understandable* (35/40, 88%), *I would find Ada flexible to interact with* (35/40, 88%), *It would be easy for me to become skillful at using Ada* (34/40, 85%), and *I would find Ada easy to use* (36/40, 90%). The mean score for these 7 questions was 84.6%. Free-text comments were prompted by two questions: "In your own words, what was MOST helpful about Ada?" and "In your own words, what was the biggest problem in using Ada?" The results are summarized in Table 6.

Table 4. Results of the user survey on previous use of technology and information seeking behavior model (N=40).

Question type, question, and response options	Participants, n (%)
Questions on use of technology and looking for medical information	
Question 1: Do you use a cellphone to send SMS text messages?	
No	1 (3)
Sometimes	4 (10)
Often	35 (88)
Question 2: Do you use a cellphone to enter or view information?	
No	3 (8)
Sometimes	7 (18)
Often	30 (75)
Question 3: Do you use a computer at work?	
No	19 (48)
Sometimes	3 (8)
Often	18 (45)
Question 4: Where would you MOST OFTEN look for medical advice?	
Doctor (physician)	22 (55)
Pharmacist	1 (3)
Family	1 (3)
Friend	0 (0)
On the web	16 (40)
Other	0 (0)
Question 5: Do you search for medical symptoms on the web?	
No	7 (18)
Sometimes	19 (48)
Often	14 (35)
Question 6: Have you used a diagnosis program or symptom checker before?	
No	29 (73)
Sometimes	10 (25)
Often	1 (3)

Table 5. Results of the user survey questions derived from the Technology Acceptance Model (N=39).

Question type, question, and response options	Participants, n (%)
Question 7: In general, how satisfied were you with Ada?	
Very unsatisfied	0 (0)
Unsatisfied	12 (31)
Neutral	14 (36)
Satisfied	13 (33)
Very satisfied	0 (0)
Top 2	13 (33)
Question 8: How enjoyable did you find using Ada?	
Very unpleasant	0 (0)
Unpleasant	16 (40)
Neutral	15 (38)
Enjoyable	9 (23)
Very enjoyable	0 (0)
Top 2	9 (23)
Question 9: Were you expecting it to be different than it was?	
No	31 (78)
Yes	9 (23)
Question 10: If a friend were in need of similar help, how likely would you be to recommend Ada to them?	
Very unlikely	1 (3)
Unlikely	2 (5)
Neutral	9 (23)
Likely	13 (33)
Very likely	15 (38)
Top 2	28 (70)
Question 11: Using Ada would enable me to record my medical symptoms and problems quickly.	
Strongly disagree	1 (3)
Disagree	1 (3)
Neutral	7 (18)
Agree	21 (53)
Strongly agree	12 (30)
Top 2	33 (83)
Question 12: Learning to use Ada would be easy for me.	
Strongly disagree	0 (0)
Disagree	2 (5)
Neutral	1 (3)
Agree	24 (60)
Strongly agree	13 (33)
Top 2	37 (93)
Question 13: I would find it easy to get Ada to do what I want it to do.	
Strongly disagree	0 (0)
Disagree	2 (5)
Neutral	12 (30)

Question type, question, and response options	Participants, n (%)
Agree	20 (50)
Strongly agree	6 (15)
Top 2	26 (65)
Question 14: The way to use Ada was clear and understandable.	
Strongly disagree	0 (0)
Disagree	2 (5)
Neutral	3 (8)
Agree	21 (53)
Strongly agree	14 (35)
Top 2	35 (88)
Question 15: I would find Ada flexible to interact with.	
Strongly disagree	0 (0)
Disagree	0 (0)
Neutral	5 (13)
Agree	26 (65)
Strongly agree	9 (23)
Top 2	35 (88)
Question 16: It would be easy for me to become skillful at using Ada.	
Strongly disagree	0 (0)
Disagree	2 (5)
Neutral	4 (10)
Agree	22 (55)
Strongly agree	12 (30)
Top 2	34 (85)
Question 17: I would find Ada easy to use.	
Strongly disagree	0 (0)
Disagree	0 (0)
Neutral	4 (10)
Agree	24 (60)
Strongly agree	12 (30)
Top 2	36 (90)

Table 6. Summary of free-text comments from the survey (N=40).

Question and responses	Participants, n (%)
In your own words, what was MOST helpful about Ada?	
Described the system as easy to use or understand [a]	14 (35)
Referred to good questions or history taking	16 (40)
No data or said "no comment" or similar	5 (13)
Other comments, including "took mind off pain" and "instant information"	5 (13)
In your own words, what was the biggest problem in using Ada?	
No comment	24 (60)
Difficulties with using it, mostly expressions of inexperience, including "don't like technology/apps," "inexperience," "crashed," and "initially confusing but then fairly simple."	7 (18)
Issues with questions and answers, including "not enough choices during questions," "putting in multiple symptoms," and "I have a lot of symptoms and it was hard to keep track of which one that app was asking more information on."	6 (15)
Other, including "trying to type with migraine" and "wanted the diagnosis."	2 (5)

[a]General descriptions of categories italicized.

Discussion

Principal Findings

This study provides preliminary data supporting both the feasibility and accuracy of a symptom checker app in an ED setting. To determine whether a symptom checker such as Ada is likely to be beneficial to patients and health systems, it is necessary to determine if it is (1) accurate at diagnosing patients based on their reported symptoms, (2) safe in its triage recommendations without a high level of overdiagnosis, (3) usable by a wide range of patients, and (4) able to positively influence patient decision-making in seeking appropriate care. This observational study provides insights into each of these criteria. The decision to test the Ada app, designed for home or community use, in an ED setting addresses the critical need to understand the performance of such tools for a full range of patient presentations. Determining the performance of a symptom checker in patients who are acutely ill in a community study would require a very large sample size and make the assessment of user experience challenging. This study is part of a 3-step evaluation plan covering different levels of patient acuity: (1) the ED-based study, (2) a similar study being completed in urgent primary care at Brown Medicine in Rhode Island, and (3) planned studies of app usage data in the community.

The overall performance of Ada on its top 5 diagnoses compared with the ED diagnoses was not significantly different from that of the study physicians assessed on their top 3 diagnoses. When compared based on the top 1 diagnoses, the physicians had substantially higher scores, with one being significantly more sensitive than Ada. As the physicians' diagnoses were based purely on the data collected by Ada, this suggests that the Ada algorithm could be improved in the area of ranking of diagnoses. A similar result was observed in a study of medical students and Ada diagnosing 3 case vignettes in rheumatology. Ada's performance was lower on the top 1 diagnosis but almost the same on the top 5 [34]. When the Ada diagnoses were critiqued by the physicians, 15% (19/130) were considered *unreasonable*, and <1% (1/130) were considered *very unreasonable*. The percentage of agreement between pairs of physicians on the final diagnoses was higher than their level of agreement with Ada, but a larger study would be required to determine if this was significant.

Scores for *comprehensiveness* and *relevance* were low in this study because of the presence of 3 to 5 diagnoses in the differential list. The relevance of the physicians might increase if they were not required to record 3 diagnoses for each case. Overall, the performance of Ada was very similar to the physicians' mean scores and matched their comprehensiveness if the top 5 diagnoses were included, but patients may have difficulty interpreting 5 options, especially as this means that the most accurate diagnosis could be one of the last two. Of note, Ada's correct diagnoses were nearly all the same as those of physician 1, suggesting that, where symptom data were adequate, the performance was good. The availability of vital sign data had little effect on the physicians' differential diagnoses or triage.

At least two independent emergency physicians rated Ada's triage recommendations as safe in 86% (32/37) of the patients. Although none of the remaining 14% (5/37) of patients whose triage recommendations were scored as *too risky* experienced an adverse outcome in the ED, the study was not powered to detect uncommon or rare serious conditions or provide longer follow-up. When considering how to improve the safety of triage recommendations of symptom checkers, weight should be given not only to the seriousness of the *most likely* diagnoses but also to the *riskiness* of certain clusters of symptoms that may represent a less common but serious condition. Many of the patients studied underwent evaluation to rule out serious conditions such as myocardial infarction. A negative evaluation does not mean that the evaluation and level of care were inherently incorrect. For example, in this study, a diagnosis of AMI by Ada or a physician using the data collected by Ada was

considered correct if the patient was screened for AMI in the ED, even if the screen was negative.

We are not aware of a study of patients directly entering data on their own symptoms to evaluate the accuracy of both diagnosis and triage in a general ED population. A study in 2 Canadian EDs and 13 primary care practices evaluated a symptom checker developed by the team that recommended a triage level out of 4 options [35]. For 281 hospital patients, the sensitivity for emergencies was 10/10 (100%) and, for urgent cases, it was 73/81 (90%), but performance for routine and home care was poorer at 52% and 29%, respectively. The positive predictive value was 40% for allocation to the hospital, 93% for primary care, and 32% for home care. They had to exclude 50% of the patients because of lack of access to the visit record and 22% of the remainder as the patients entered a different presenting complaint. The triage performance of Ada and other symptom checkers can be compared with telephone or nurse-based triage systems. A report showed a median accuracy of 75% and potential harm from undertriage of 1.3% to 3.2% [36], although there was wide variation in the reported performance in that review. Tam et al [37] reviewed a range of studies of triage in the ED and primary care in 2018, including designs using case vignettes or retrospective chart reviews. They reported that "when comparing with all multi-center studies, both methods revealed a triage accuracy of about 60% and about 23% of cases [it] was under-estimated," although some single-site studies had higher accuracies of >70%.

The user survey showed that the patients had a wide range of ages (up to 76 years, with 12/40, 30% being aged ≥50 years); a broad range of levels of schooling; and a wide range of racial, ethnic, and socioeconomic backgrounds (Table 1). Although they had mixed views on their overall satisfaction with Ada, this was likely their first use of the app and, for 73% (29/40), their first use of any symptom checker. They had generally positive views on the system and would recommend it to others. This was in the context of patients typically feeling very unwell and in a stressful ED environment, providing some confidence that patients can use a well-designed symptom checker even when sick. Satisfaction scores would presumably be higher if the patients had been given access to their diagnoses and triage results, as noted in some patient comments (eg, "wanted the diagnosis").

A previous study of Ada that assessed its ease of use in a primary care setting in the United Kingdom [28] showed that younger patients were more likely to report that Ada provided helpful advice, but there was no effect seen from patients' sex. A study in California [18] also examined the mix of patients using Ada in a health system and found that the patient characteristics of users were similar to those of their general patient population but with a younger mean age. A study by Knitza et al [38] on the use of Ada, a custom diagnostic app (Rheport), and web-based searching for symptoms, showed a high System Usability Scale score of 77.1 for Rheport and a somewhat lower score for Ada of 74.4 (P<.001). Ratings for "very helpful" or "helpful" were higher for Rheport (65.8% vs 44.3%), although similar numbers would recommend each system to others (79.2% vs 73.3%).

Limitations

Some technical problems occurred with the research environment for Ada developed for the study (not with the public-facing production system), which led to missing or incomplete reports in some cases. These problems were addressed in the follow-up study.

The greatest challenge encountered in this study design was defining the "correct diagnosis" based on the clinicians' assessment and EHR notes from the ED visit. In many cases, the ED physicians' role was to exclude serious causes for the presenting symptoms, with the patient potentially seeking investigation through their primary care physician or specialist at a later date. The multiple possible diagnoses recorded in the ED physicians' notes also make it difficult to compare metrics on matching diagnoses with many previous studies of symptom checkers. In the studies by Semigran et al [12] and Gilbert et al [13], clinical vignettes had just 1 correct diagnosis, which, as shown here, is not typical for actual patient assessments in the ED. The *comprehensiveness* and *relevance* metrics consider the full set of diagnoses in each differential. In the large vignette study by Gilbert et al [13], *comprehensiveness* for Ada was similar to that in this study (48% on the top 3 diagnoses vs 41% in this study), but relevance was higher (45% on the top 3 diagnoses vs 22% in this study); Ada's performance improved in this study when the top 5 diagnoses were included.

In addition, in those cases where a comparison was able to be made with the ED physicians' diagnoses, >27% (8/30) were not correctly diagnosed by either Ada or the study physicians. This is likely due in part to the lack of data on physical examinations and investigations available to the ED physician. Berry et al [39] compared the diagnostic accuracy of (1) physicians reviewing the symptom data items collected by the symptom checker (similar to this study) with (2) adding the clinical data collected by the reviewing physician. They confirmed that the additional clinical data significantly improved diagnostic accuracy. The use of the symptom data from Ada potentially limits the diagnostic accuracy of the physician reviewers as they cannot ask additional questions. This may lead to underestimating the physician performance compared with Ada's.

We collected data in 2019 before the COVID-19 pandemic and, therefore, did not include patients with COVID-19. Initial studies of symptom checkers on clinical case descriptions of patients with COVID-19 have shown a wide range of performance for different symptom checkers on different data sets. The follow-up study is expected to include some patients with COVID-19.

Comparison With Other Studies

Semigran et al [12] studied 23 symptom checkers using 45 clinical vignettes. The best-performing symptom checkers had a sensitivity for the correct diagnosis for their top 1 diagnosis in 35.5% of cases and for the top 3 diagnoses in 51% of cases, compared with independent physicians, who achieved 72% top-1 sensitivity and 84.3% top-3 sensitivity [23]. The 3 best-performing systems were closer in performance to the physicians, with top-1 sensitivity of 43% to 50% and top-3

sensitivity of 67% to 71%. A recent repeat study used the vignettes by Semigran et al [12] and versions of symptom checker apps available in 2020 [40]. Top-1 sensitivity was 10% higher at 45.5%, and top-10 sensitivity was 71.1% (compared with 55.8% for the top 20 diagnoses in the study by Semigran et al [12]). Top-1 sensitivity for Ada was higher than the average at 53%, with the top-10 sensitivity average at 71%. Another study that sought to replicate the study by Semigran et al [12] was published by Hill et al [24] in 2020 with symptom checker apps available in Australia. In that study, the mean sensitivity for the correct diagnoses was 36% for the top 1, 52% for the top 3, and 58% for the top 10. In our study, sensitivity based on the top 3 diagnoses was 63.3% for Ada versus 70% for physicians, with a larger difference including only the top 1 diagnoses. However, the results in our study are not completely comparable given that many cases had more than one ED diagnosis (potentially increasing apparent performance) and were based on data entry by patients of their own symptoms (potentially decreasing performance). A recent study on the use of Ada for mental health disorders by Henneman et al [29] used a similar study design and metric to our RIH study, with 49 patients (61% female) using the app before seeing a psychotherapist. As in our RIH study, they used the metric for a correct diagnosis as *at least one diagnosis from the Ada differential diagnosis matching with one from the psychotherapist*. They showed a correct match (sensitivity) of the Ada top 1 diagnosis in 51% (95% CI 37.5%-64.4%) of cases and with Ada's top 5 diagnoses in 69% (95% CI 55.4%-80.6%) of cases, a higher performance than in our RIH study for the top 1 diagnosis but equivalent for the top 5 diagnoses. The interrater reliability varied widely depending on the condition.

In a systematic review, Chambers et al [7] included all available evaluation studies of symptom checkers used for the assessment of urgent conditions up to April 2018. A total of 27 studies were included in the final review. They identified the potential risks associated with symptom checker use as "increasing demand," "duplicating healthcare contacts," and "providing advice that is not safe or clinically appropriate." They found "little evidence to indicate whether or not digital and online symptom checkers are detrimental to patient safety." Limitations in the 6 studies that measured safety included being short-term; having small samples and, therefore, including insufficient adverse events; and being limited to specific symptoms or from specific populations unrepresentative of urgent care users. Among the priorities for research, they identified qualitative research to investigate perceptions of symptom checkers and barriers to their use by people who are less familiar with digital technology. Our study provides a model for the assessment of safety and usability based on direct use by patients with serious and potentially life-threatening conditions, but this sample size does not allow for a clear assessment of triage accuracy and safety. In addition, an intervention study would be required to show whether changes in patients' decisions to seek urgent care based on the symptom checker output affected the quality and safety of care. More recent studies by our group have a larger sample size, and planned studies will address the impact on patient decision-making.

Previous studies of symptom checkers in urgent or ED settings have generally been limited by the selection of less urgent patients—for example, Mediktor [41]—or lack of direct use of symptom checkers by patients. Berry et al [39] studied WebMD, iTriage, and FreeMD symptom checkers for the diagnosis of cough in primary care. The best-performing symptom checker had a sensitivity for the correct diagnosis of 34.5% (top 1 diagnosis) and 71.6% (top 3 diagnoses). However, the symptom data came from paper forms patients filled out in the waiting room, not from direct use of an app; moreover, the study had a small list of possible diagnoses, and the "correct diagnosis" was allocated by one primary care physician, limiting the potential generalization of the findings. A team at the Queen Mary Hospital in Hong Kong studied the triage accuracy of 2 symptom checkers based on a random sample of 100 charts from the ED [42]. Triage accuracy was rated as low at 74% and 50%, with poorer performance on more urgent cases; diagnostic accuracy was not tested. A recent study showed that patients reading simplified case vignettes based on the study by Semigran et al [12] had better triage performance than symptom checkers on low-risk conditions but poorer performance in detecting cases requiring urgent care [43]. This suggests that there may be an important role for symptom checkers in identifying serious conditions missed by patients, but that only the best-performing symptom checkers are likely to be effective.

Developments in symptom checker algorithms in the last 7 years would have been expected to improve triage performance (in line with improvements in diagnosis) but, as shown by Schmieding et al [40], there are still many symptom checkers that are less safe and more prone to overtriage and possibly undertriage than patients themselves. In the aforementioned 2020 study by Schmieding et al [40] using the protocol and vignettes by Semigran et al [12] (with small modifications), the median triage accuracy was 55.8%, similar to the 2015 study by Semigran et al [12] (59.1%), but the ratio of over- to undertriage shifted, with overtriage only slightly higher than undertriage. This led to the apps missing >40% of the emergency vignettes overall, a real concern for patient safety. The Ada app performed better than average with a triage accuracy of 64%, which is close to the results of our RIH study, and a high accuracy for triage of emergencies (89%). The study by Hill et al [24], which also replicated the study by Semigran et al [12], showed a mean triage accuracy of 49%, with stronger performance on emergency cases. The study by Chan et al [35] (reported earlier) showed significantly better triage accuracy by their locally developed symptom checker than the patients overall (73% vs 58%; $P<.01$), and performance on emergency and urgent cases was stronger than on routine or home care. A recent study on the use of Ada by 378 "walk-in" patients in urgent care compared its triage accuracy with the result of a triage nurse using the Manchester Triage System [44]. The app was shown to undertriage 8.9% of cases and overtriage 57.1%, although physician assessment of the undertriaged cases suggested that 14 cases (3.7%) did not represent a risk of an adverse event. Overtriage was significantly higher than that in our RIH study. The results are not fully comparable as all triage performances in our RIH study were judged by 3 physicians rather than by a triage nurse.

Potential Change in Patient Decision-making

There is initial evidence that patients may change the urgency level of the care they seek based on the results of symptom checkers and other web-based diagnostic tools. A study of 158,000 patient consultations with the symptom checker Buoy evaluated the intended urgency level of the care they would seek before and after viewing the output of the symptom checker. A total of 32% of patients stated that they would seek a lower urgency level, and 4% would seek a higher level [26]. A study of Ada in use by patients in a primary care practice in the United Kingdom evaluated any change in their intended acuity of care based on the diagnosis and triage results. A total of 12.8% of the patients stated that they would seek less urgent care, and 1.2% said that they would seek more urgent care [28]. Therefore, it is likely that a proportion of patients using symptom checkers will change their care-seeking behavior based on the results presented. The main limitation of these 2 studies is that there is no indication of whether the advice received was accurate or safe. Some evidence of the benefits of web-based health data was shown in a study of 5000 laypeople who were asked to review clinical vignettes of different illnesses and then provide their assessment of the appropriate triage level and top 3 diagnoses [45]. After viewing web-based information about the case (most commonly search engines and specialist medical sites), the participants' diagnostic accuracy improved modestly from 49.8% to 54% (P<.001). There was no change in triage accuracy. A similar study by Martin et al [46] compared the ability of patients with low-acuity symptoms in an ED to match at least two of the differential diagnoses made by the physician who subsequently assessed them. A total of 300 patients were randomized to (1) receive assistance from a standard Google search, (2) receive assistance from a Google search with enhanced medical features, or (3) have no access to searching. There was no significant difference in the percentage matching 2 physician diagnoses (27%, 28.3%, and 23.8%, respectively). Given the potential for patients to seek less urgent care based on symptom checker assessment results and the limited evidence of improved diagnostic accuracy from widely used search engines, studies of diagnosis and triage accuracy of symptom checkers are essential. In addition, studies are required on the effects of symptom checker output on patients' care-seeking behavior and the safety and appropriateness of those decisions.

Conclusions

The primary goal of this study was to answer whether a widely used symptom checker was safe, effective, and usable by patients who were acutely ill and might have a life-threatening disease. These data, in the context of existing studies of symptom checker apps, including Ada, should help provide validation of diagnosis and triage accuracy with real patient use. This pilot study provides evidence to support usability and on overall diagnostic performance while showing the potential for improving the ranking of diagnoses by Ada. On triage, performance was similar to that of the clinicians in most cases but with significant overtriage and some undertriage. A larger study would be required to provide definitive evidence, and assess the potential impacts on care-seeking behavior. The results also demonstrate a fundamental challenge in developing and evaluating such systems—gaps and variability in the documentation of differential diagnoses in EHRs. A larger study is underway of patients requesting urgent primary care appointments, with patients completing the consent form, the Ada questions, and the user survey in the home or community setting. In addition to broadening the range of diagnoses and patient types, this will allow for a better assessment of appropriate triage levels. The goal is also to use the research environment and data to study a range of symptom checkers. The design of this study should provide a model for larger and more varied evaluation studies of real-world performance and use of symptom checkers. Good performance in observational studies of this sort is a requirement for measuring the likely clinical impact in intervention studies, including randomized controlled trials.

Acknowledgments

The authors would like to thank the staff at Rhode Island Hospital for their help with this study. They would also like to thank Ada Health for providing a modified version of the Ada symptom checker app that supported the study protocol, advice on the clinical use of the system, and assistance with Figure 1.

Conflicts of Interest

No funding for this study was provided by Ada Health, but support was provided in adapting the Ada app for the research study, which was necessary for this research design. The design, implementation, analysis, and write-up of the study and the decision to publish were made by the authors at Brown University without influence from Ada Health. The authors did agree to allow Ada research staff to review the study for factual accuracy regarding the Ada software. Minor updates were made to the explanation of the missing case reports, and some additional data were added to Table 3 to provide context.

Multimedia Appendix 1
Example Ada case output.
[PDF File (Adobe PDF File), 68 KB - mhealth_v10i9e38364_app1.pdf]

Multimedia Appendix 2
Table of all study cases, including presenting complaints and emergency department diagnoses.

[XLSX File (Microsoft Excel File), 19 KB - mhealth_v10i9e38364_app2.xlsx]

Multimedia Appendix 3
Diagnoses in cases where Ada urgency or triage advice was rated as unsafe or too risky.
[DOCX File , 20 KB - mhealth_v10i9e38364_app3.docx]

References

1. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
2. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. JAMA 1987 Jul 03;258(1):67-74. [doi: 10.1001/jama.258.1.67] [Medline: 3295316]
3. Vardell E, Moore M. Isabel, a clinical decision support system. Med Ref Serv Q 2011;30(2):158-166. [doi: 10.1080/02763869.2011.562800] [Medline: 21534115]
4. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii33-ii39 [FREE Full text] [doi: 10.1136/bmjqs-2012-001623] [Medline: 23893394]
5. Udayaraj UP, Haynes R, Winearls CG. Late presentation of patients with end-stage renal disease for renal replacement therapy--is it always avoidable? Nephrol Dial Transplant 2011 Nov;26(11):3646-3651. [doi: 10.1093/ndt/gfr164] [Medline: 21454353]
6. Hamilton W. Diagnosis: cancer diagnosis in UK primary care. Nat Rev Clin Oncol 2012 Apr 03;9(5):251-252. [doi: 10.1038/nrclinonc.2012.63] [Medline: 22473095]
7. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
8. Thompson M. Online symptom checkers: evolution of self diagnosis is upon us. Save a Life by NHCPS. 2016 Jul 6. URL: https://nhcps.com/online-symptom-checkers-evolution-of-self-diagnosis-is-upon-us/ [accessed 2022-08-26]
9. WebMD symptom checker with body map. WebMD. URL: https://symptoms.webmd.com/default.htm#/info [accessed 2022-07-28]
10. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
11. Ćirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
12. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
13. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
14. When to use 111. NHS. URL: https://www.nhs.uk/using-the-nhs/nhs-services/urgent-and-emergency-care/nhs-111/ [accessed 2022-07-28]
15. Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. Arthritis Res Ther 2021 Apr 13;23(1):112 [FREE Full text] [doi: 10.1186/s13075-021-02498-8] [Medline: 33849654]
16. Symptom checker. Mayo Clinic. URL: https://www.mayoclinic.org/symptom-checker/select-symptom/itt-20009075 [accessed 2022-01-20]
17. Babylon health homepage. Babylon Health. URL: http://babylonhealth.com/en-us [accessed 2022-08-26]
18. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
19. Turner J, O'Cathain A, Knowles E, Nicholl J. Impact of the urgent care telephone service NHS 111 pilot sites: a controlled before and after study. BMJ Open 2013 Nov 14;3(11):e003451 [FREE Full text] [doi: 10.1136/bmjopen-2013-003451] [Medline: 24231457]
20. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. Health Aff (Millwood) 2017 Mar 01;36(3):485-491. [doi: 10.1377/hlthaff.2016.1130] [Medline: 28264950]
21. Mansab F, Bhatti S, Goyal D. Performance of national COVID-19 'symptom checkers': a comparative case simulation study. BMJ Health Care Inform 2021 Mar;28(1):e100187 [FREE Full text] [doi: 10.1136/bmjhci-2020-100187] [Medline: 33685943]

22. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
23. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. JAMA Intern Med 2016 Dec 01;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]
24. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
25. K health homepage. K health. URL: https://khealth.com/symptom-checker/ [accessed 2022-07-28]
26. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
27. Health. Powered by Ada. Ada. URL: https://ada.com/ [accessed 2022-07-28]
28. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
29. Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022 Jan 31;9(1):e32832 [FREE Full text] [doi: 10.2196/32832] [Medline: 35099395]
30. Raman S, Fraser HS. Evaluation of Diagnostic Apps for Myocardial Infarction and Other Causes of Chest Pain: Informing Patient Use. Evaluation of Diagnostic Apps for Myocardial Infarction and Other Causes of Chest Pain: Society for Improvement of Medical Diagnosis; 2021 Presented at: 14th Annual International Conference on Diagnostic Error in Medicine; 10/25/2021; Online URL: https://repository.library.brown.edu/studio/item/bdr:52u24yfa
31. 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]
32. Shemesh T, Barnoy S. Assessment of the intention to use mobile health applications using a technology acceptance model in an israeli adult population. Telemed J E Health 2020 Sep;26(9):1141-1149. [doi: 10.1089/tmj.2019.0144] [Medline: 31930955]
33. Fraser HS, Long WJ, Naimi S. Comparing complex diagnoses: a formative evaluation of the heart disease program. Proc AMIA Annu Fall Symp 1997:853. [doi: 10.1016/0076-6879(90)86140-q] [Medline: 2233313]
34. Knitza J, Tascilar K, Gruber E, Kaletta H, Hagen M, Liphardt A, et al. Accuracy and usability of a diagnostic decision support system in the diagnosis of three representative rheumatic diseases: a randomized controlled trial among medical students. Arthritis Res Ther 2021 Sep 06;23(1):233 [FREE Full text] [doi: 10.1186/s13075-021-02616-6] [Medline: 34488887]
35. Chan F, Lai S, Pieterman M, Richardson L, Singh A, Peters J, et al. Performance of a new symptom checker in patient triage: Canadian cohort study. PLoS One 2021;16(12):e0260696 [FREE Full text] [doi: 10.1371/journal.pone.0260696] [Medline: 34852016]
36. Blank L, Coster J, O'Cathain A, Knowles E, Tosh J, Turner J, et al. The appropriateness of, and compliance with, telephone triage decisions: a systematic review and narrative synthesis. J Adv Nurs 2012 Dec;68(12):2610-2621. [doi: 10.1111/j.1365-2648.2012.06052.x] [Medline: 22676805]
37. Tam HL, Chung SF, Lou CK. A review of triage accuracy and future direction. BMC Emerg Med 2018 Dec 20;18(1):58 [FREE Full text] [doi: 10.1186/s12873-018-0215-0] [Medline: 30572841]
38. Knitza J, Muehlensiepen F, Ignatyev Y, Fuchs F, Mohn J, Simon D, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. Front Public Health 2022;10:1 [FREE Full text] [doi: 10.3389/fpubh.2022.844669] [Medline: 35273944]
39. Berry AC, Berry NA, Wang B, Mulekar MS, Melvin A, Battiola RJ, et al. Symptom checkers versus doctors: a prospective, head-to-head comparison for cough. Clin Respir J 2020 Apr;14(4):413-415. [doi: 10.1111/crj.13135] [Medline: 31860762]
40. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
41. Arancibia JN, Sanchez FM, Mejías AD, Castillo G, Vilaplana C, García Briñon M, et al. Evaluation of a diagnostic decision support system for the triage of patients in a hospital emergency department. Int J Interact Multimed Artif Intell 2019;5(4):60. [doi: 10.9781/ijimai.2018.04.006]
42. Yu SW, Ma A, Tsang VH, Chung LS, Leung S, Leung L. Triage accuracy of online symptom checkers for accident and emergency department patients. Hong Kong J Emergency Med 2019 Apr 16;27(4):217-222. [doi: 10.1177/1024907919842486]
43. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]

44. Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, et al. Safety of triage self-assessment using a symptom assessment app for walk-in patients in the emergency care setting: observational prospective cross-sectional study. JMIR Mhealth Uhealth 2022 Mar 28;10(3):e32340 [FREE Full text] [doi: 10.2196/32340] [Medline: 35343909]
45. Levine DM, Mehrotra A. Assessment of diagnosis and triage in validated case vignettes among nonphysicians before and after internet search. JAMA Netw Open 2021 Mar 01;4(3):e213287 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3287] [Medline: 33779741]
46. Martin SS, Quaye E, Schultz S, Fashanu OE, Wang J, Saheed MO, et al. A randomized controlled trial of online symptom searching to inform patient generated differential diagnoses. NPJ Digit Med 2019 Nov 11;2(1):110 [FREE Full text] [doi: 10.1038/s41746-019-0183-0] [Medline: 31728417]

Abbreviations

AMI: acute myocardial infarction
ED: emergency department
EHR: electronic health record
REDCap: Research Electronic Data Capture
RIH: Rhode Island Hospital

Edited by L Buis; submitted 02.04.22; peer-reviewed by J Knitza, R Riessen, M Schmieding; comments to author 27.04.22; revised version received 31.05.22; accepted 10.06.22; published 19.09.22.

Please cite as:
Fraser HSF, Cohan G, Koehler C, Anderson J, Lawrence A, Pateña J, Bacher I, Ranney ML
Evaluation of Diagnostic and Triage Accuracy and Usability of a Symptom Checker in an Emergency Department: Observational Study
JMIR Mhealth Uhealth 2022;10(9):e38364
URL: https://mhealth.jmir.org/2022/9/e38364
doi:10.2196/38364
PMID:36121688

Original Paper

The Effects of a Lifestyle Intervention Supported by the InterWalk Smartphone App on Increasing Physical Activity Among Persons With Type 2 Diabetes: Parallel-Group, Randomized Trial

Ida Kær Thorsen[1], PhD; Yanxiang Yang[2], MSc; Laura Staun Valentiner[3], MSc; Charlotte Glümer[4], PhD; Kristian Karstoft[1,5], PhD; Jan Christian Brønd[6], PhD; Rasmus Oestergaard Nielsen[7,8], PhD; Charlotte Brøns[9,10], PhD; Robin Christensen[11,12], PhD; Jens Steen Nielsen[13], PhD; Allan Arthur Vaag[10], PhD; Bente Klarlund Pedersen[1], PhD; Henning Langberg[3], PhD; Mathias Ried-Larsen[1], PhD

[1]Center of Inflammation and Metabolism and Centre for Physical Activity Research, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

[2]Chair of Sport and Health Management, Technical University of Munich, Munich, Germany

[3]CopenRehab, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[4]Centre for Diabetes, Municipality of Copenhagen, Copenhagen, Denmark

[5]Department of Clinical Pharmacology, Bispebjerg Hospital, Copenhagen, Denmark

[6]Research Unit for Exercise Epidemiology, Centre of Research in Childhood Health, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[7]Department of Public Health, Aarhus University, Aarhus, Denmark

[8]Research Unit for General Practice, Aarhus University, Aarhus, Denmark

[9]Department of Endocrinology, Diabetes and Bone-metabolic Research Unit, Rigshospitalet, Copenhagen, Denmark

[10]Steno Diabetes Center Copenhagen, Gentofte, Denmark

[11]Section for Biostatistics and Evidence-Based Research, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

[12]Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[13]Danish Centre for Strategic Research in Type 2 Diabetes, Steno Diabetes Center Odense, Odense University Hospital, Odense, Denmark

Corresponding Author:
Ida Kær Thorsen, PhD
Center of Inflammation and Metabolism and Centre for Physical Activity Research
Copenhagen University Hospital - Rigshospitalet
Blegdamsvej 9
Copenhagen, 2100
Denmark
Phone: 45 28700785
Email: ida.kaer.thorsen@regionh.dk

Abstract

Background: Effective and sustainable implementation of physical activity (PA) in type 2 diabetes (T2D) health care has in general not been successful. Efficacious and contemporary approaches to support PA adherence and adoption are required.

Objective: The primary objective of this study was to investigate the effectiveness of including an app-based (InterWalk) approach in municipality-based rehabilitation to increase moderate-and-vigorous PA (MVPA) across 52 weeks compared with standard care among individuals with T2D.

Methods: The study was designed as a parallel-group, randomized trial with 52 weeks' intervention and subsequent follow-up for effectiveness (52 weeks from baseline). Participants were recruited between January 2015 and December 2016 and randomly allocated (2:1) into 12 weeks of (1) standard care + InterWalk app–based interval walking training (IWT; IWT group; n=140), or (2) standard care + the standard exercise program (StC group; n=74). Following 12 weeks, the IWT group was encouraged to maintain InterWalk app–based IWT (3 times per week for 30-60 minutes) and the StC group was encouraged to maintain exercise without structured support. Moreover, half of the IWT group (IWTsupport group, n=54) received additional motivational support following the 12-week program until 52-week follow-up. The primary outcome was change in objectively measured MVPA time

(minutes/day) from baseline to 52-week follow-up. Key secondary outcomes included changes in self-rated physical and mental health–related quality of life (HRQoL), physical fitness, weight, and waist circumference.

Results: Participants had a mean age of 59.6 (SD 10.6) years and 128/214 (59.8%) were men. No changes in MVPA time were observed from baseline to 52-week follow-up in the StC and IWT groups (least squares means [95% CI] 0.6 [–4.6 to 5.8] and –0.2 [–3.8 to 3.3], respectively) and no differences were observed between the groups (mean difference [95% CI] –0.8 [–8.1 to 6.4] minutes/day; *P*=.82). Physical HRQoL increased by a mean of 4.3 (95% CI 1.8 to 6.9) 12-item Short-Form Health Survey (SF-12) points more in the IWT group compared with the StC group (Benjamini-Hochberg adjusted *P*=.007) and waist circumference apparently decreased a mean of –2.3 (95% CI –4.1 to –0.4) cm more in the IWT group compared with the StC group but with a Benjamini-Hochberg adjusted *P*=.06. No between-group differences were observed among the remaining key secondary outcomes.

Conclusions: Among individuals with T2D referred to municipality-based lifestyle programs, randomization to InterWalk app–based IWT did not increase objectively measured MVPA time over 52 weeks compared with standard health care, although apparent benefits were observed for physical HRQoL.

Trial Registration: ClinicalTrials.gov NCT02341690; https://clinicaltrials.gov/ct2/show/NCT02341690

(JMIR Mhealth Uhealth 2022;10(9):e30602) doi:10.2196/30602

KEYWORDS

type 2 diabetes mellitus; exercise; telemedicine; primary health care; accelerometry; quality of life; waist circumference; mHealth; mobile app

Introduction

Physical activity (PA) is a cornerstone in the prevention and management of type 2 diabetes (T2D) [1] and adults with T2D are recommended to perform a minimum of 150 minutes of moderate-and-vigorous PA (MVPA) per week [2]. However, effective and sustainable implementation of PA programs in health care has in general not been successful [3-7]. These barriers may include low self-efficacy, inappropriate goal-setting, lack of access to facilities, and lack of supervision and social support [2]. Previous efforts to support increased PA levels among individuals with T2D have produced promising results in trial contexts, but the extensive support applied may involve limited longer-term effectiveness of these efforts [8-12]. Accordingly, sustained increases in PA levels are rarely reported [10-15]. Exercise supervision may be required to improve glycemic control among individuals with T2D, whereas advice alone is insufficient [16]. Altogether, continued support to sustain PA levels is pivotal. Given the growing global prevalence and incidence of T2D [17], providing fully supervised exercise on a life-long basis is unfeasible. Thus, efficacious and contemporary approaches to support PA adherence are required.

While MVPA is recommended for individuals with T2D [1,2], brief high-intensity exercise bouts may also be effective in increasing physical fitness and improving glycemic control [18]. Accordingly, we have previously shown that 4 months of technology-supported interval walking training (IWT; 5 sessions of 60 minutes/week) led to increased physical fitness, decreased body mass and adiposity, and improved glycemic control among individuals with T2D, whereas energy expenditure–matched continuous walking did not [19]. In addition to promoting increased peak intensities of PA, IWT is a safe and convenient exercise type that has proven effective in maintaining adherence and motivation to continue IWT after a trial, especially when receiving feedback from a training device [20].

The increasing implementation of digital solutions in health care [1,21], along with the growing smartphone ownership among the older populations (≥50 years) [22], suggests the use of smartphones as an easy, accessible opportunity for remote and flexible PA support. Accordingly, the InterWalk app for smartphones was developed in Denmark to deliver individually tailored IWT as a feasible intervention to promote PA among individuals with T2D [20,23]. Further, we have previously observed that motivational support consisting of goal setting, SMS text message support, and phone calls increased adherence to InterWalk app–based IWT [24].

The primary aim of this study was to test the hypothesis that InterWalk app–based IWT implemented in a municipality-based health care setting is superior in increasing MVPA across 52 weeks compared with standard care among individuals with T2D. Secondarily, we investigated the effects of the intervention on self-rated physical and mental health–related quality of life (HRQoL), physical fitness, weight, and waist circumference. Moreover, we wanted to explore the effects of additional motivational support for InterWalk app–based IWT on these outcomes.

Methods

Study Design

The study was a parallel-group, randomized trial with 52 weeks of intervention and subsequent effectiveness follow-up a year after baseline. All participants provided oral and written informed consent prior to commencing any study procedures. The original protocol has been published [23], while the prespecified statistical analysis plan is available as Multimedia Appendix 1. Reporting is in accordance with the CONSORT statement.

Ethics Approval

The study was approved by the scientific ethical committee of the Capital Region of Denmark (H-1-2014-074) and registered at clinicaltrials.gov (NCT02341690).

Participants and Eligibility

Inclusion criteria were T2D diagnosis, ≥18 years of age, and referral to a municipality health promotion center or hospital in the participating municipality by the individual's general practitioner. Exclusion criteria were medical contraindications to exercise, for example, chronic complications in the musculoskeletal system, painful osteoarthritis, or heart conditions [25]; declining to participate in an exercise program at the health promotion center or hospital; current participation in other intervention studies; or insufficient Danish language skills. All individuals who met for an appointment at the participating health promotion centers or hospital were screened through medical records and at a screening interview with a health professional at the center or hospital.

The data were collected at the participating health promotion centers (Amager, Vanløse, Østerbro, Vesterbro in the Municipality of Copenhagen, Denmark; and Municipality of Guldborgsund, Denmark) and a hospital (Bornholm, Denmark).

Randomization and Blinding

Participants were randomly allocated (2:1) into 1 of 2 arms: (1) standard care + InterWalk app–based IWT (IWT group) or (2) standard care + the standard exercise program (StC group). Following the initial 12-week supervised exercise program, the participants in the IWT group were further randomly allocated (1:1) into (1) IWT, no additional support (IWT_{only} group) or (2) IWT, with additional support ($IWT_{support}$ group), that is, participants in the IWT_{only} group and the $IWT_{support}$ group underwent similar interventions during the 12-week exercise program, and allocation to either of these was concealed until after the 12-week intermediate assessment.

Participants were randomized using random permuted blocks stratified by sex (2 levels) and center (6 levels). The allocation sequence was generated through a standardized computer program by an independent statistician (RC) and stored on a password-protected computer by an independent researcher (RN) who was not involved in any study procedures. Following the completion of all baseline measurements, the independent researcher was contacted and performed allocation (to StC or IWT_{only} or $IWT_{support}$). The respective group allocation for the initial 12 weeks (StC or IWT) was subsequently returned by email to the health professional who informed the participant about the allocation by telephone call. Information about IWT_{only} or $IWT_{support}$ was not disclosed by the independent researcher to the health professionals until week 12. Following the 12-week intermediate assessment, the independent researcher was contacted and the allocation into IWT_{only} or $IWT_{support}$ was returned by email to the health professional who informed the participant about the allocation by telephone call. The health professionals carried through the data collection and intervention, and thus, were only blinded to the primary outcome.

Interventions

Development and Implementation

The interventions have been described in detail elsewhere [23]. Briefly, the interventions were designed to comply with the standard health care in Denmark [26]. Study investigators (LV, CB, and HL) prepared and led several workshops totally 16 hours), where the health professionals and the investigators discussed the normal work routines in detail and discussed suggestions on how to implement the InterWalk app and co-interventions into the daily routine. Based on these discussions, the intervention protocol was developed. Following the finalization of the study protocol, the health professionals completed an educational program (15 hours in total), where they were trained in implementation of the study procedures and manuals. In addition, the health care professionals attended workshops every second month throughout the trial period to ensure the procedures and manuals were consistently implemented as described in the protocol.

Baseline to 12-Week Follow-up

As part of normal practice, lifestyle interventions may be prescribed to individuals with T2D by their general practitioner by referral to municipality-based lifestyle programs [26]. This program entails a combination of exercise, diabetes education on self-management, smoking cessation courses, and diet counseling. A patient might receive all or any combination of these components, depending on the specific need of the patient. The decision is based on a dialog between the patient and the health care provider upon initiation of the program.

Both the IWT and StC groups underwent 12 weeks of standard municipality-based health care [23,26]. They both received an exercise program during the initial 12 weeks. The StC group was prescribed a standard municipality-based exercise program (2 sessions per week for 12 weeks; combined aerobic and resistance training delivered by trained health professionals). The IWT group was prescribed InterWalk app–based IWT (30-60 minutes per session, 3 sessions per week for 12 weeks; for an extended description, see below) instead of the standard exercise program. During the 12-week exercise program, IWT was group based and 2 of the 3 sessions were supervised.

From 12-Week to 52-Week Follow-up

Following the initial 12 weeks, the StC group was encouraged to maintain exercise without structured support.

The participants allocated to the InterWalk app–based intervention from baseline to 12-week follow-up were either allocated to be encouraged to maintain InterWalk app–based IWT (3 times per week for 30-60 minutes; IWT_{only}) until 52-week follow-up or was, for explorative purposes, allocated to additional motivational support following the 12-week program and until the 52-week follow-up ($IWT_{support}$). Feasibility and usability have been described elsewhere [24]. The motivational support included (1) individual motivational interviews with individual goal setting performed by the health care professionals, (2) InterWalk-based IWT with voluntary ambassadors affiliated with the Danish Diabetes Association

(a Danish not-for-profit patient organization) once per week, and (3) SMS text message support (once per week).

The motivational interviews were semistructured and performed during weeks 16, 20, 28, and 40 after the baseline assessment. They were designed to structure the communication between the participant and the health professional. The interviews intended to facilitate a partnership formed to reveal and visualize the patients' motivation and barriers toward the intended behavior change and acknowledgement of patient autonomy [27]. Moreover, the health care professionals were instructed to help the participants perform goals toward a lifestyle change (ie, not only increased PA) using the S.M.A.R.T. principle derived from the Goal Setting Theory [23]. In the S.M.A.R.T. principle S denotes *specific*, M denotes *measurable* (eg, can we track it?), A denotes *achievable* (eg, is it realistic to obtain?), R is for *relevant* (eg, does it make sense for the participant?), and T denotes *timely* (eg, is it obtainable within a prespecified period?).

The ambassador program was a part of a peer-to-peer educational program, where the Danish Diabetes Association offered group walking for all members. The participants were invited to attend these walking groups, where the walking activity was implemented using the InterWalk app.

Finally, facilitation of high adherence to IWT was based on automated feedback to self-reported adherence to IWT using weekly bidirectional SMS text message surveys [24]. An automatic survey was sent during the afternoon on Sundays and inquired about the frequency of IWT during the past week (1="I have walked more than 3 times/week", 2="I have walked 3 times/week", 3="I have walked 2 times/week", 4="I have walked 1 times/week", and 5="I did not walk"). If no answer was received, a reminder was sent within 24 hours. If no answer was received following the reminder or the participant answered, "I did not walk," the health care professional was instructed to reach the participant by phone using a semistructured approach. The semistructured interview guide consisted of 2 overall questions: (1) Do you experience barriers toward IWT? If yes, which ones? (2) How can I help you to overcome these barriers? If the participant indicated walking 1-2 times/week, an automated SMS text message was sent encouraging the participant to walk more.

The InterWalk App

The design and functionality of the InterWalk app was developed in collaboration with persons with T2D using a participatory design. The full details have been described in detail elsewhere [20]. Briefly, following the initial testing of the exercise modality [19], we developed a mock-up of the app and invited persons with T2D to provide their initial feedback. Following the development of the beta-version of the app, 3 iterations of user feedback were incorporated into the design before releasing the version used in this study (versions 8 and 9) [20].

The app was designed to specifically facilitate IWT through continuous individualized audio feedback. The app guides and paces the user through repeated cycles of 3-minute slow walking and 3-minute fast walking [20]. IWT was individualized based on a 7-minute standardized walking test in the InterWalk app [20], which the participants performed at baseline and were asked to repeat every 4 weeks. The intensity was derived from the onboard accelerometer. The individualized cut-offs were determined during the 7-minute walking test and implemented in the IWT sessions. The cut-offs were based on the median intensity between moderate and fast walking pace (the lower limit of intensity during fast IWT walking) and low and moderate walking pace (the upper limit of intensity during slow IWT walking) during the 7-minute walking tests. If participants exceeded these cut-offs during IWT (below the lower limit during fast walking or above the upper limit during slow walking), they received an audio prompt encouraging them to increase or decrease the walking pace. Following each IWT session, the app displayed data from the session performed, such as information about walking distance, steps, duration, and the quality of performance. The latter was determined using the fraction of training time (percentage) which the participant spent within the intensity cut-offs.

Following an IWT session, the objectively measured adherence (intensity, duration, and frequency) to IWT sessions was calculated based on the electronically logged information from the InterWalk app. Data from the InterWalk app were continuously uploaded to a central and secure server throughout the duration of the trial. Self-reported adherence to IWT was calculated based on data from the SMS text message survey, including frequency of weekly use of the InterWalk app and reasons for not using the InterWalk app [24].

Outcomes

Primary and key secondary outcome measurements are described in further detail in the published protocol [23]. Outcome measurements were conducted by trained health professionals at the respective center or hospital.

The primary outcome measure was change in objectively measured MVPA time (minutes/day) from baseline to 52-week follow-up. PA outcomes were assessed using accelerometers (AX3; Axivity) attached to the participants' thigh and back using an adhesive tape (Fixomull Stretch; BSN Medical Inc.) and worn for 7-10 consecutive days at baseline, 12 weeks, and 52 weeks. Accelerometer setup and download of raw data (100 Hz) and downsampling to 30 Hz were performed using OmGui (version 1.0.0.28) [28]. The data were then exported to ActiGraph raw binary files (gt3x files) and resampled into counts (agd files) using ActiLife (version 6.11.6). The final data reduction and generation of PA outcome variables were done using a custom-built software developed at the University of Southern Denmark [29]. MVPA time was defined according to the Freedson cut point, ≥1952 counts per minute (CPM) [30], using the vertical axis of the accelerometer placed on the back. Participants were included in the analyses of objectively measured PA if they had ≥3 days of ≥22 hours of measurement (ie, allowing for a 2-hour nonwear time [31]). Accelerometer wear time and nonwear time are presented in Multimedia Appendix 2.

Key secondary outcome measures include changes in physical and mental HRQoL, physical fitness (peak oxygen consumption [VO_{2peak}]), self-rated PA energy expenditure (PAEE), exercise

motivation, weight, and waist circumference, which were assessed at baseline, 12 weeks, and 52 weeks. Physical and mental HRQoL were assessed using the 12-item Short-Form Health Survey (SF-12), including 8 subscales [32]. The Physical Component Summary (PCS; score 0-100) was calculated based on the 4 subscales: Physical Functioning, Role Physical, Bodily Pain, and General Health; and the Mental Component Summary (MCS; score 0-100) was calculated based on the 4 subscales: Vitality, Social Functioning, Role Emotional, and Mental Health [32]. VO_{2peak} (ml O_2/minute) was estimated by regression based on the participants' weight, height, sex, and the acceleration (G) during the last 30 seconds of fast walking in the 7-minute standardized walking test in the InterWalk app [33]. Self-rated PAEE was assessed using the Recent Physical Activity Questionnaire (RPAQ) [34]. Exercise motivation was assessed using the Relative Autonomy Index (RAI) for the Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2) [35]. Weight (kg) was measured using an electronic weight, and waist circumference (cm) was measured midway between the most distal part of the costae and the most proximal part of the iliac crest.

Exploratory secondary outcomes were changes in objectively measured light-intensity PA (100 to <1952 CPM) time (minutes/day), total PA level (CPM), and steps (numbers/day) [30,36] assessed by accelerometers worn on the participants' back as described above. Sitting time (minutes/day) was derived from the accelerometer placed on the thigh as described elsewhere [31]. BMI (kg/m^2) was calculated based on weight and height. Any adverse events were reported by the participants to the health professionals at the center or hospital.

Sample Size

The minimal important difference was considered to be 10 minutes of MVPA per day. Based on existing experimental evidence, we assumed the SD of the change in MVPA time from baseline to 52-week follow-up to be between 1.2 and 2.3 times the difference between groups [37]; accordingly, SD twice the minimal important difference (20 minutes of MVPA per day) was used in the sample size calculations. To maintain a conservative sample size, considering multiple comparisons in a 3-arm trial, Bonferroni adjustment was used in the power calculation. A total of 190 participants were required to achieve a statistical power (1-β) of 80% with an α level of .017 (0.5/3) using an unpaired *t* test (2-sided). Allowing for 30% attrition, 272 participants (91 in the StC group and 181 in the IWT group) should be recruited. The intervention settings enabled recruitment until December 15, 2016, and thus the sample size would include 272 participants or truncated at the number of participants included at the end of the recruitment period—whichever was reached first. On December 15, 2016, 214 participants had been included and they constituted the final analysis population.

Statistical Methods

A statistical analysis plan was developed and published at the Centre for Physical Activity Research website prior to commencing the statistical analyses [38]; see Multimedia Appendix 1. All continuous outcomes were analyzed using repeated-measures mixed linear models with the baseline value of the relevant variable as a covariate, including fixed effect factors for group 1 (2 levels: IWT/StC), group 2 (3 levels: IWT_{only}, $IWT_{support}$, and StC), time (3 levels: baseline, 12-week, 52-week), and group 1 × time interaction; and random effects (patient ID). The analyses were adjusted for the stratifying factors sex (2 levels: male and female) and center (6 levels: Amager, Vanløse, Østerbro, Vesterbro, Guldborgsund, and Bornholm). The following model assumptions were investigated: (1) linearity, (2) normality of residuals, (3) homogeneity of residuals variance, and (4) independence of residual error. Data were analyzed based on the intention-to-treat population (no imputation for missing data) as observed, including all randomized participants with available data at baseline; repeated measures mixed linear models were valid assuming that data were "missing at random." Results are reported as least squares means (LS means) for each group and the difference between them with 95% CIs. Analyses of categorical outcomes for dichotomous endpoints were performed using the Fisher exact test and reported based on the observed proportions.

To investigate the robustness of the primary analyses, 5 sensitivity analyses were conducted, including 4 analyses of differences in changes in MVPA time from baseline to 52-week follow-up between the IWT group and the StC group: (1) per-protocol analysis, including complete cases of the primary outcome at baseline and 52-week follow-up with ≥70% of the prescribed exercise (3 sessions per week for 52 weeks) completed in the IWT group and no registered IWT sessions in the InterWalk app in the StC group; (2) analysis including all randomized participants with available data at baseline with a conservative single-imputation nonresponder imputation technique (ie, missing data replaced with the baseline observation carried forward); (3) analysis with the subsample of participants not undergoing intervention during a period of InterWalk app malfunctions connected to a major restructuring of the iOS (version 9; available during the period from September 16 to October 21, 2016); and (4) analysis with daytime criteria for inclusion, that is, ≥3 days of ≥14 hours of daytime measurement (6 AM to 22 PM; ie, allowing for 2 hours of nonwear time during this period). In the fifth sensitivity analysis, we used an alternative cut-off at 3000 CPM to investigate changes in forced walking time from baseline to 52-week follow-up between the IWT group and the StC group.

Subgroup analyses were conducted to investigate the effects of IWT versus StC on change in MVPA time (minutes/day) after 52 weeks among subgroups of sex (men/women); T2D duration (≤5 years/>5 years); alcohol consumption (within recommended levels/above recommendations); smoking habits (smoker/nonsmoker); highest level of education (International Standard Classification of Education 2011 [ISCED-2011] levels 0-4/ISCED-2011 levels 5-8); civil status (single, divorced, or widowed/married or cohabiting); baseline SF-12 PCS level (high/low); and baseline SF-12 MCS level (high/low). SF-12 PCS and MCS were divided into high and low using median split. Moreover, differences in median (25th to 75th percentile) duration of IWT (minutes/week) over 52 weeks were investigated among these subgroups.

All reported 95% CIs and *P* values are 2-sided. Because more than 1 comparison was made, key secondary outcome analyses were controlled for the false discovery rate using the Benjamini–Hochberg step-up procedure [39]. All statistical analyses were performed using Stata/IC 13 (StataCorp).

Results

Overview

Between January 2015 and December 2016, 762 individuals were screened for inclusion (Figure 1). Of these, 548 were excluded primarily due to declining to participate (n=209) or physical challenges (n=100). Of the 214 participants enrolled in the study, 74 participants were allocated to the StC group and 140 participants to the IWT group. A total of 130 participants completed the 52-week follow-up, which was finalized in December 2017. Participant demographic and clinical characteristics are presented in Table 1. The mean age of participants was 59.6 (SD 10.6) years and 128/214 (59.8%) were men. They had a mean BMI of 34.6 (SD 6.6) kg/m^2, a median (25th-75th percentile) T2D duration of 2.6 (0.2-8.6) years (Table 1). At baseline, participants spent a median (25th-75th percentile) of 30.2 (16.9-48.1) minutes/week in MVPA (Table 1).

The IWT group engaged in IWT with a median (25th-75th percentile) frequency of 0.3 (0.1-0.7) sessions/week for 11.4 (4.1-22.3) minutes/week at a median intensity of 0.17 (0.11-0.23) G (Multimedia Appendices 3 and 4). For the subgroup that completed the 7-minute standardized walking test in the InterWalk app at baseline, this corresponded to an average intensity across sessions of 65% (SD 49-78) of their maximal intensity at baseline. According to self-reported IWT adherence, participants in the IWT$_{support}$ group used the InterWalk app 1-2 times/week or more in most weeks; the primary reason for not using the InterWalk app was illness (Multimedia Appendix 4). For the IWT$_{support}$ group, 148/188 (78.7%) motivational interviews were completed during follow-up (Multimedia Appendix 5). However, only 27/47 (57%) participants completed all 4 interviews and the adherence to the Danish Diabetes Association walking groups was low as only 5-6/47 (11%-12%) participants participated in this activity. Only a limited number of participants included IWT as a specific part of the goals during follow-up (18/47, 38%).

Figure 1. Participant flowchart.

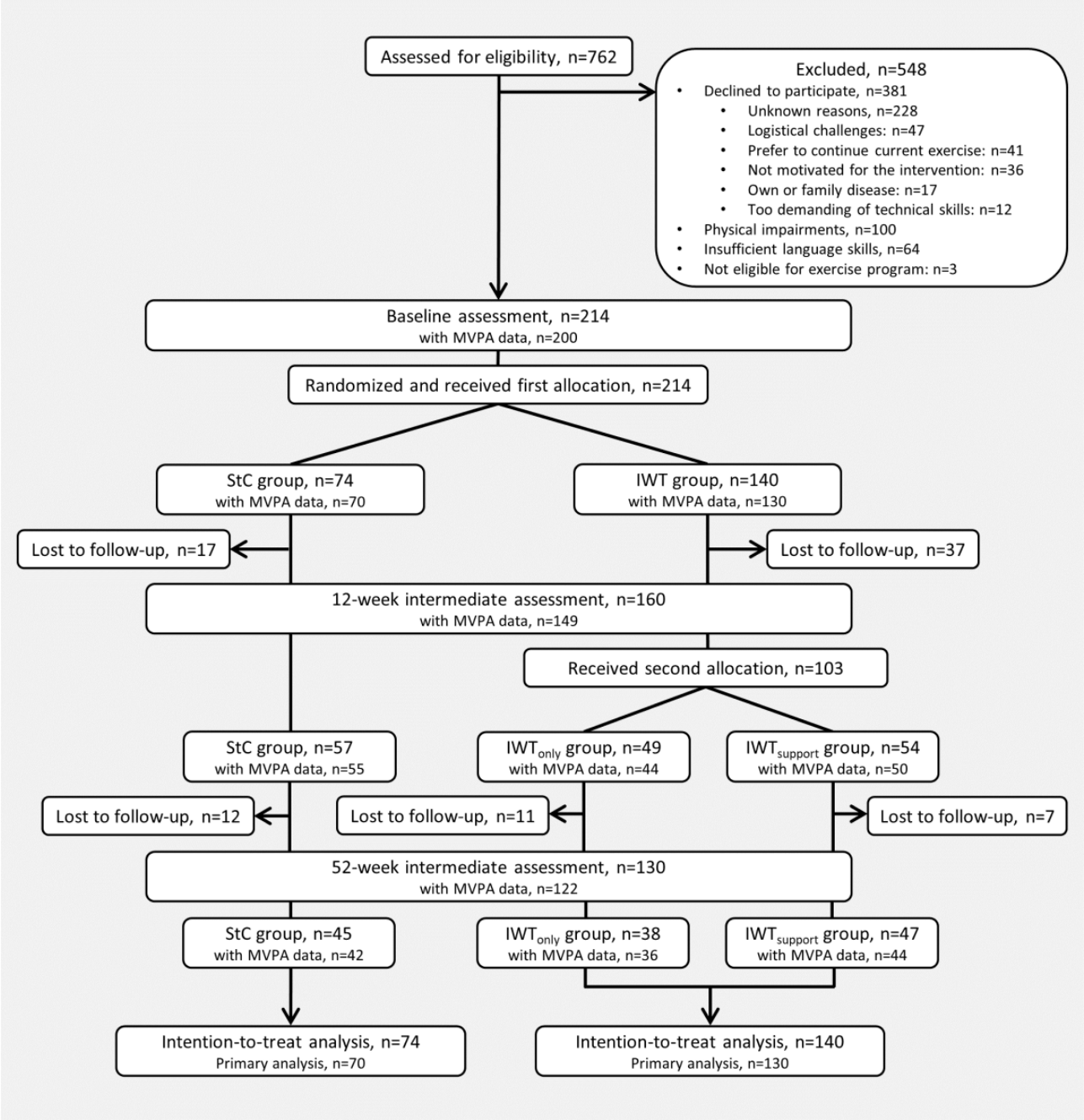


Table 1. Demographic and clinical characteristics of participants at baseline (n=214 patients in the intention-to-treat population).[a]

Demographic and clinical characteristics	StC[b] group (n=74)	IWT[c] group (n=140)	Total (n=214)
Sex			
Male	47 (63.5)	81 (57.9)	128 (59.8)
Female	27 (36.5)	59 (42.1)	86 (40.2)
Age (years)	59.8 (10.1)	59.6 (10.8)	59.6 (10.6)
Type 2 diabetes duration (n=197)	1.7 (0.2-7.0)	3.0 (0.2-10.0)	2.6 (0.2-8.6)
≤5 years	45/68 (66.2)	74/129 (57.4)	119 (60.4)
>5 years	23/68 (33.8)	55/129 (42.6)	78 (39.6)
Alcohol consumption (n=213)			
Within the recommended levels	69 (93.2)	129/139 (92.8)	198 (93.0)
Above recommendations	5 (6.8)	10/139 (7.2)	15 (7.0)
Smoking habits			
Smoker	20 (27.0)	29 (20.7)	49 (22.9)
Nonsmoker	54 (73.0)	111 (79.3)	165 (77.1)
Highest level of education (n=209)			
ISCED-2011[d] levels 0-4	39 (52.7)	75/135 (55.6)	114 (54.5)
ISCED-2011 levels 5-8	35 (47.3)	60/135 (44.4)	95 (45.5)
Civil status			
Single, divorced, or widowed	36 (48.6)	64 (45.7)	100 (46.7)
Married or cohabiting	38 (51.4)	76 (54.3)	114 (53.3)
Height (cm; n=212)	171.7 (9.4)	172.5 (8.8)	172.2 (9.0)
Physical activity and fitness			
MVPA[e] time (minutes/day; n=200)	30.5 (17.3-41.7)	30.2 (16.6-51.8)	30.2 (16.9-48.1)
Sitting time (minutes/day; n=195)	556.3 (157.5)	527.3 (162.6)	537.4 (161.0)
LPA[f] time (minutes/day; n=200)	136.3 (52.5)	136.0 (48.7)	136.1 (49.9)
TPA[g] level (CPM[h]; n=200)	212.9 (141.6-291.0)	207.6 (137.8-345.2)	211.5 (140.9-313.8)
Steps (n/day; n=200)	4164 (2766-5738)	4078 (2801-6145)	4100 (2794-5986)
VO_{2peak}[i] (ml O_2/minute; n=124)	1859 (511)	1763 (469)	1791 (481)
Self-reported measures			
SF-12[j] Physical Component Summary (score 0-100)	40.3 (10.4)	41.5 (10.1)	41.1 (10.2)
SF-12 Mental Component Summary (score 0-100)	47.4 (39.6-55.8)	50.9 (41.3-57.1)	50.1 (40.1-56.9)
RPAQ[k] self-rated PAEE[l] (kJ/kg/day; n=209)	150.1 (92.5-226.7)	136.0 (92.9-225.7)	143.8 (92.9-226.4)
BREQ-2[m] RAI[n] (score –24 to 20)	7.4 (4.0-10.8)	7.4 (3.5-11.7)	7.4 (3.5-11.3)
Body composition			
Weight (kg; n=212)	101 (85-119)	98 (87-113)	99 (87-115)
Waist circumference (cm; n=211)	117.9 (16.0)	115.3 (12.9)	116.2 (14.1)
BMI (kg/m^2; n=212)	34.6 (6.6)	33.6 (5.3)	34.0 (5.7)

[a]Data are presented as median (IQR), mean (SD), n (%), or n/N (%).

[b]StC: standard care.

[c]IWT: interval walking training.
[d]ISCED-2011: International Standard Classification of Education 2011.
[e]MVPA: moderate-and-vigorous physical activity.
[f]LPA: light-intensity physical activity.
[g]TPA: total physical activity.
[h]CPM: counts per minute.
[i]VO_{2peak}: peak oxygen consumption.
[j]SF-12: 12-item Short-Form Health Survey.
[k]RPAQ: Recent Physical Activity Questionnaire.
[l]PAEE: physical activity energy expenditure.
[m]BREQ-2: Behavioral Regulation in Exercise Questionnaire-2.
[n]RAI: Relative Autonomy Index.

Change in Moderate-and-Vigorous Physical Activity Time

No change in MVPA time was observed from baseline to 52-week follow-up within the StC and IWT groups (*P*=.81 and *P*=.91, respectively; Figure 2) and no between-group difference was observed (–0.8 minutes/day; 95% CI –8.1 to 6.4 minutes/day; *P*=.82; Table 2). From baseline to 12-week follow-up, MVPA increased by a mean of 3.6 (95% CI 0.2 to 6.9) minutes/day in the IWT group and was unchanged in the StC group. However, no difference was observed between the groups (Multimedia Appendix 6).

Only 5 participants in the IWT group fulfilled the per-protocol criteria and the analysis was thus omitted. The remaining sensitivity analyses supported the primary analysis (Multimedia Appendix 7).

Figure 2. Least squares means (95% CI's) of MVPA time (min/day) at baseline, 12-week and 52-week follow-up for the IWT and StC group. IWT: interval walking training; MVPA: moderate-and-vigorous physical activity; StC: standard care.

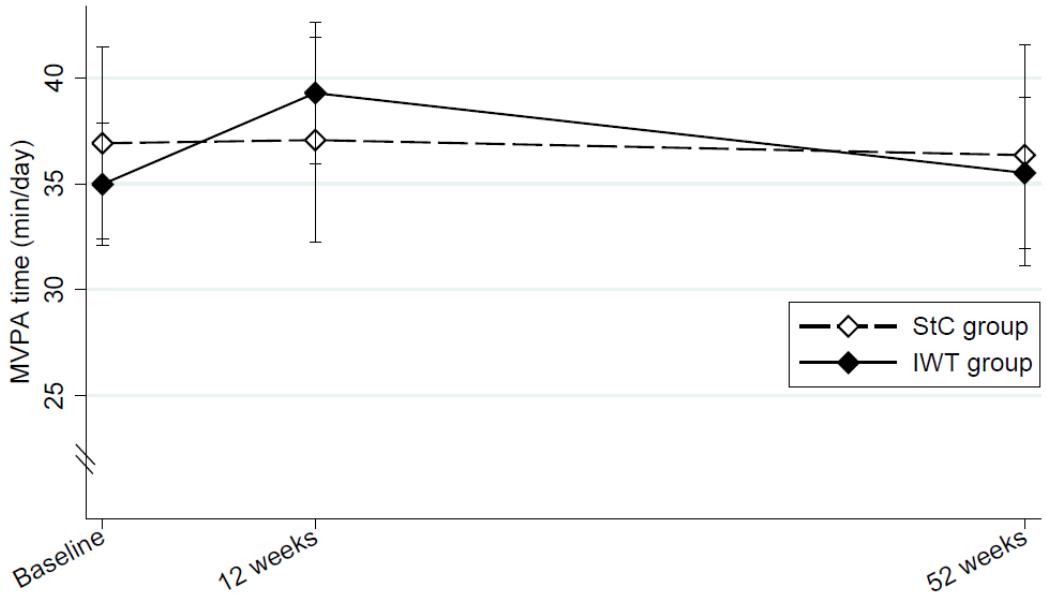


Table 2. Intention-to-treat analyses of changes from baseline to 52-week follow-up between the IWT group and the StC group[a].

Outcomes	StC[b] group (n=74)	IWT[c] group (n=140)	Between-group difference		
			Difference between means (95% CI)	*P* value	BH[d] adjusted *P* value
Primary outcome					
MVPA[e] time (minutes/day; n=200)	0.6 (–4.6 to 5.8)	–0.2 (–3.8 to 3.3)	–0.8 (–8.1 to 6.4)	.82	N/A[f]
MVPA responders[g] (n=116)	7/41 (17.1)	17/75 (22.7)	N/A	NA[h]	N/A
Key secondary outcomes					
SF-12[i] PCS[j] (score 0-100)	0.0 (–1.9 to 1.8)	4.3 (3.1 to 5.6)	4.3 (1.8 to 6.9)	.001	.007
SF-12 PCS responders[g] (n=128)	15/45 (33.3)	45/83 (54.2)	N/A	NA	N/A
SF-12 MCS[k] (score 0-100)	2.6 (0.2 to 5.0)	1.3 (–0.4 to 2.9)	–1.3 (–4.7 to 2.0)	.43	.43
SF-12 MCS responders[g] (n=128)	18/45 (40.0)	31/83 (37.3)	N/A	NA	N/A
VO_{2peak}[l] (ml O_2/min; n=54)	4.0 (–128.0 to 136.0)	76.3 (5.4 to 147.3)	72.4 (–102.0 to 246.8)	.42	.49
RPAQ[m] self-rated PAEE[n] (kJ/kg/day; n=209)	8.0 (–15.9 to 31.8)	–8.3 (–24.3 to 7.8)	–16.2 (–49.2 to 16.7)	.33	.46
BREQ-2[o] RAI[p] (score –24 to 20)	1.3 (–0.3 to 2.8)	2.6 (1.6 to 3.6)	1.4 (–0.7 to 3.5)	.20	.35
Weight (kg; n=212)	–0.1 (–1.3 to 1.1)	–1.6 (–2.3 to –0.8)	–1.4 (–3.1 to 0.2)	.09	.21
Waist circumference (cm; n=212)	–0.9 (–2.2 to 0.5)	–3.1 (–4.0 to –2.2)	–2.3 (–4.1 to –0.4)	.02	.06
Exploratory secondary outcomes					
Sitting time (minutes/day; n=195)	39.5 (–2.4 to 81.4)	–13.0 (–41.8 to 15.8)	–52.5 (–109.6 to 4.6)	NA	N/A
LPA[q] time (minutes/day; n=200)	–15.9 (–25.6 to –6.2)	–2.9 (–9.6 to 3.7)	13.0 (–0.5 to 26.4)	NA	N/A
TPA[r] level (CPM[s]; n=200)	–12.9 (–41.6 to 15.9)	3.9 (–15.6 to 23.4)	16.8 (–23.1 to 56.6)	NA	N/A
Steps (n/day; n=200)	–439 (–1100 to 222)	168 (–280 to 616)	607 (–311 to 1525)	NA	N/A
BMI (kg/m^2; n=212)	–0.1 (–0.5 to 0.4)	–0.5 (–0.8 to –0.3)	–0.5 (–1.0 to 0.1)	NA	N/A

[a]Data for the StC and IWT groups are presented as least squares mean (95% CI) or n (%) or n/N (%).

[b]StC: standard care.

[c]IWT: interval walking training.

[d]BH: Benjamini-Hochberg.

[e]MVPA: moderate-and-vigorous physical activity.

[f]N/A: not applicable.

[g]MVPA responders were defined as change in MVPA time ≥10 minutes/day; SF-12 PCS responders were defined as change in SF-12 PCS >3.29 [40]; SF-12 MCS responders were defined as change in SF-12 MCS >3.77 [40].

[h]NA: not analyzed.

[i]SF-12: 12-item Short-Form Health Survey.

[j]PCS: Physical Component Summary.

[k]MCS: Mental Component Summary.

[l]VO_{2peak}: peak oxygen consumption.

[m]RPAQ: Recent Physical Activity Questionnaire.

[n]PAEE: physical activity energy expenditure.

[o]BREQ-2: Behavioral Regulation in Exercise Questionnaire-2.

[p]RAI: Relative Autonomy Index.

[q]LPA: light-intensity physical activity.

[r]TPA: total physical activity.

[s]CPM: counts per minute.

Key Secondary Outcomes

From baseline to 12- and 52-week follow-up, SF-12 PCS score (ie, physical HRQoL) increased by a mean of 4.3 (95% CI 3.1 to 5.6) points and 4.2 (95% CI 3.0 to 5.4) points, respectively, in the IWT group and remained unchanged in the StC group. Thus, the SF-12 PCS score increased by a mean of 3.7 (95% CI 1.2 to 6.1) points and 4.3 (95% CI 1.8 to 6.9) points (Benjamini–Hochberg adjusted *P*=.007) more in the IWT group compared with the StC group at 12- and 52-week follow-up, respectively (Multimedia Appendix 6 and Table 2). According to post hoc linear regression analysis, a change in MVPA and IWT duration was not associated with a change in SF-12 PCS over 52 weeks (Multimedia Appendix 8).

Waist circumference decreased by a mean of –3.1 (95% CI –4.0 to –2.2) cm from baseline to 52-week follow-up in the IWT group with no change in the StC group (Table 2). Thus, waist circumference apparently decreased by a mean of –2.3 (95% CI –4.1 to –0.4) cm more in the IWT group compared with the StC group (Benjamini-Hochberg adjusted *P*=.06; Table 2). From baseline to 12-week follow-up, waist circumference decreased by a mean of –2.9 (95% CI –3.7 to –2.0) cm and –3.2 (95% CI –4.5 to –1.9) cm in the IWT and StC group, respectively. No differences were observed between the groups (Multimedia Appendix 6). According to a post hoc linear regression analysis, every 10 minute/day increase in MVPA was associated with a mean decrease of 0.6 (95% CI –1.1 to –0.1) cm in waist circumference (r=–0.21; *P*=.03), while every 10 minute/week increase in IWT duration was associated with a mean decrease of 0.6 (–1.1 to –0.2) cm in waist circumference (r=–0.30; *P*=.008; Multimedia Appendix 8).

No differences in the changes in SF-12 MCS (mental HRQoL), VO_{2peak}, RPAQ self-rated PAEE, BREQ-2 RAI, and weight from baseline to 12- and 52-week follow-up, respectively, were observed between the groups (Multimedia Appendix 6 and Table 2).

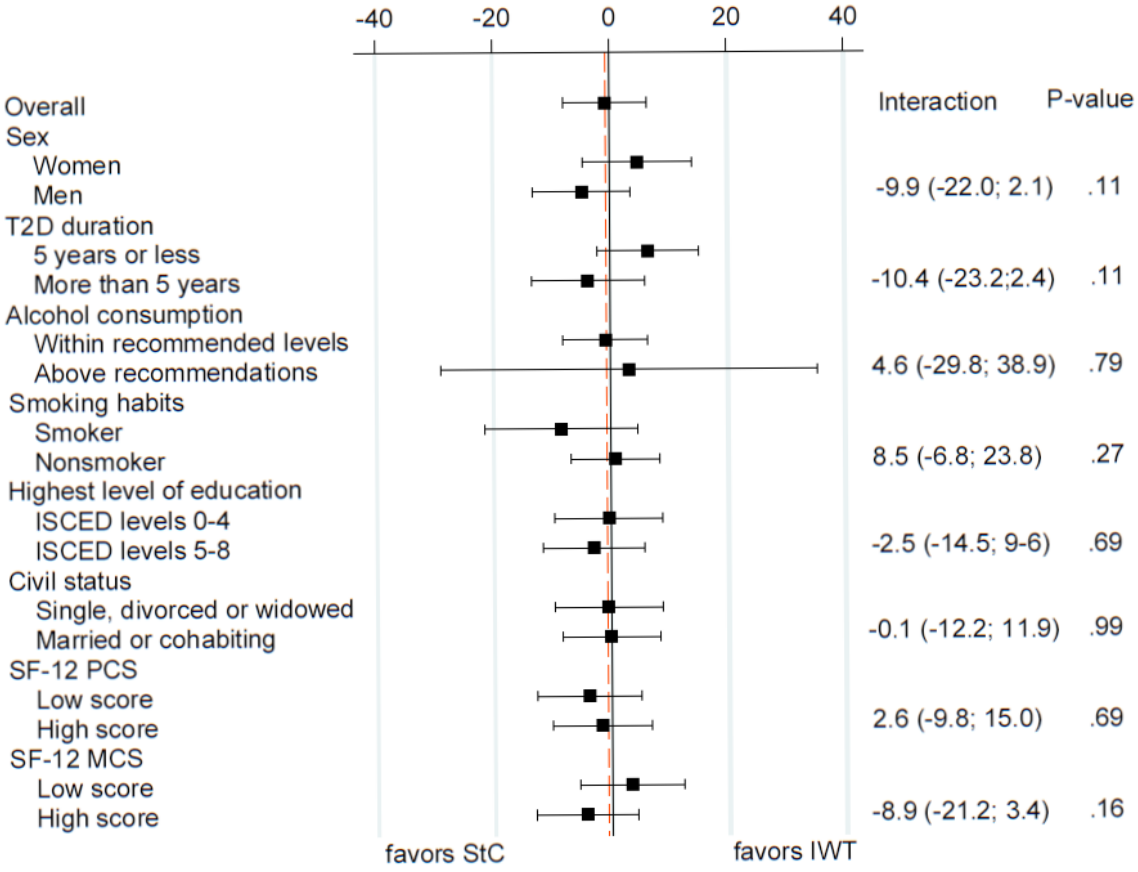
Exploratory Secondary Outcomes

Among the exploratory secondary outcomes, no between-group differences were observed from baseline to 12- and 52-week follow-up, except for light-intensity physical activity time and steps that increased from baseline to 12-week follow-up by a mean of 20.1 (95% CI 7.5 to 32.7) minutes/day and 1124 (95% CI 255 to 1992) steps/day, respectively, more in the IWT group compared with the StC group (Multimedia Appendix 6 and Table 2).

Subgroup Analyses

From baseline to 52-week follow-up, there were no differences between the IWT_{only} and $IWT_{support}$ groups in the changes in primary, key secondary, and exploratory secondary outcomes (Multimedia Appendix 9). Moreover, no subgroup effects of IWT versus StC on the change in MVPA time were observed (*P*>.1 for interaction; Figure 3), nor were there any differences in mean IWT duration over 52 weeks among the subgroups in the IWT group (Multimedia Appendix 10).

Figure 3. Forest plot of overall and subgroup effects of IWT vs. StC on change in MVPA time (min/day) after 52 weeks. Subgroup effects include sex (men/women), type 2 diabetes duration (≤5 years/>5 years), alcohol consumption (within the recommended levels/above recommendations), smoking habits (smoker/nonsmoker), highest level of education (ISCED-2011 levels 0-4/ISCED-2011 levels 5-8), civil status (single, divorced or widowed/married or cohabiting), SF-12 PCS (high/low), and SF-12 MCS (high/low), assessed at baseline. ISCED: International Standard Classification of Education; IWT: interval walking training; MCS: Mental Component Summary; PCS: Physical Component Summary; SF-12: 12-item Short-Form Health Survey; StC: standard care; T2D: type 2 diabetes.



Harms

No adverse events or harms were reported to the health professionals.

Discussion

The main finding of this study was that InterWalk app–based IWT did not increase objectively measured MVPA time over 52 weeks compared with standard care among individuals with T2D referred to municipality-based lifestyle programs. While InterWalk app–based IWT resulted in improvements in physical HRQoL and nonsignificant reduction in waist circumference compared with StC, no effects were observed in mental HRQoL, physical fitness, or weight. A key objective of this study was to support individuals with T2D in maintaining PA levels after a 12-week municipality-based exercise program with minimal direct support from health professionals. In general, IWT adherence was low across 52 weeks, largely explaining the lacking effect on MVPA time. During the 12-week exercise program, IWT adherence was remarkably higher and MVPA time increased by >10% from baseline (by 3.6 minutes/day from 30.2 minutes at baseline), indicating that InterWalk app–based IWT potentially contributed to increased MVPA time. The discrepancy of the IWT adherence observed in this and previous IWT studies among individuals with T2D indicates that InterWalk app–based IWT may not—in the present circumstances—be feasible for maintaining PA level after a municipality-based exercise program [19,24].

In line with previous studies, adherence to the exercise intervention decreased over time [12] when supervision was subtracted [15]. In contrast to our observations, previous reports observed increases in objectively measured PA (such as MVPA, steps, and moderate-intensity walking time) following digitally supported PA interventions [41-44]. This discrepancy may partly be explained by shorter intervention periods of the previous studies [41-44]. Indeed, intermediate assessment after only 12 weeks in this study involved increases in MVPA time, although this effect may be driven by the additional direct supervision integral to the exercise program. As such, discrepancies may further be explained by differences in other intervention features, such as health professional consultations or group sessions in previous studies, and features of the applied digital solutions, including web-based solutions or smartphone apps [41-44]. These digital solutions were designed to support PA adoption, for example, through goal setting and self-management [41-44], whereas the InterWalk app was designed as a training device specifically to deliver individually tailored IWT [20]. However, in contrast to previous findings [41-44], providing motivational interviews with individual goal setting in addition to the

InterWalk app did not influence intervention adherence or outcomes over 52 weeks.

In line with this study, previous studies have observed decreased waist circumference [10,11,44] and improved self-reported physical health and quality of life [43,45] after a PA intervention. We observed that decreased waist circumference was maintained 40 weeks after the 12-week exercise program when other intervention features were discontinued, and participants only had access to the InterWalk app. This observation is supported by similar results in the eCoFit trial over just 10 weeks following a smartphone-supported PA intervention, featuring workout circuits using the outdoor environment, social support, goal setting, etc. [44]. Post hoc analyses suggested an inverse dose-response relationship of both IWT and MVPA with waist circumference. Accordingly, Ross and colleagues [46] observed that 24 weeks of supervised high-volume exercise decreased waist circumference among individuals with obesity. We observed similar decreases in waist circumference across 52 weeks with access to the InterWalk app and markedly less direct supervision. These results may be of clinical relevance as 1-cm decreases in waist circumference is associated with 4% reductions in visceral fat mass [47]. Thus, the observed 3-cm decrease in waist circumference is likely associated with substantial reductions in visceral fat mass, which potentially lead to considerable improvements in glycemic control and decreased low-grade inflammation [19,47,48]. Likewise, improved physical HRQoL was maintained 40 weeks after the exercise program when participants only had access to the InterWalk app. This is supported by similar results in a previous study, where physical HRQoL was preserved following a 12-month intensive lifestyle intervention [45]. Post hoc analyses suggested that the improved physical HRQoL across 52 weeks in this study was independent of changes in MVPA and IWT duration, indicating that this effect was driven by other mechanisms than PA behavior change. This is supported by a previous study observing no effects of a 1-year behavior change intervention on objectively measured PA, although higher levels of self-reported physical functioning compared with standard care were reported [49]. In this study, possible mechanisms may be related to the safe and convenient nature of IWT, potentially affecting participants' perception of their physical health, or to the continued access to app-supported PA. However, when participants in a previous study were provided access to a digital monitoring and feedback solution, the effects of self-management support on self-reported physical health were inhibited [41].

Limitations of this study include the high loss to follow-up corresponding to 39% at 52 weeks and the low adherence to the intervention. As indicated by the high adherence levels in previous IWT studies among individuals with T2D [19,24], these low adherence levels may result from the pragmatic approach to the study design. Accordingly, the applied intervention has previously showed high level of efficacy in explanatory trials including rather homogenous samples and increased standardization and control [19,24]. With our study design, we did not identify a feasible solution for effective implementation of this efficacious intervention in municipality-based health care of individuals with T2D. One limitation of our approach includes the lack of detailed information on the adherence to and delivery of the co-interventions during the municipality-based rehabilitation programs from baseline to 12-week follow-up. Besides, the lack of objective monitoring of the adherence to the follow-up intervention and the single features of the InterWalk app preclude us from drawing strong conclusions of the effectiveness of the InterWalk app per se. Moreover, according to descriptive post hoc analyses, baseline demographic and clinical characteristics among attenders and nonattenders at 52-week follow-up did not seem to be different (Multimedia Appendix 11), and thus did not appear to explain the high loss to follow-up. Further, the sample representativeness was limited. First, individuals of low socioeconomic status are less likely to be referred to municipality-based lifestyle programs [50]. Second, choosing to participate in an exercise program indicates a will to change PA behavior, whereas rejecting participation may not. Third, as Danish language was used during assessment (eg, questionnaires) and exercise programs, non-Danish speaking individuals were precluded from participation. Altogether, this sample may underrepresent the most vulnerable individuals with T2D. Further, our sample was more physically active at baseline than the general US population [51] and individuals with T2D in a previous study [52]. This indicates potential ceiling effect of MVPA at baseline and thus selection bias by exclusion of less physically active individuals from this sample. This stresses the need for development of efficacious, contemporary approaches to support PA adherence and adoption with the potential to increase PA levels among the wide population of individuals with T2D. Finally, we experienced technical malfunction of the InterWalk app related to a major restructuring of the iOS, which may have affected data upload and collection during that period. However, this was not reflected in the sensitivity analysis excluding participants undergoing intervention during this period.

In conclusion, among individuals with T2D referred to municipality-based lifestyle programs, randomization to InterWalk app–based IWT did not increase objectively measured MVPA time over 52 weeks compared with standard care, although an improvement in physical HRQoL was observed. Moreover, in this municipality-based setting, adherence to the intervention was low even when additional motivational support was provided. Further research is needed to identify optimal implementation of digital support for PA adherence and adoption among individuals with T2D.

Acknowledgments

Participants and health professionals at the centers and hospital are acknowledged for their contribution to data collection and carrying through the interventions. Cecilie Fau Brinkløv and Malte Kongstad are acknowledged for their significant contributions to the study. The InterWalk app uses the REDCap facility at OPEN, Open Patient Data Explorative Network, Odense University

Hospital, Region of Southern Denmark to collect data. The Centre for Physical Activity Research (CFAS) is supported by TrygFonden (grants IDs 101390, 20045, and 125132). During the study period, the Centre of Inflammation and Metabolism (CIM) was supported by a grant from the Danish National Research Foundation (DNRF55). CIM/CFAS is a member of DD2—the Danish Center for Strategic Research in Type 2 Diabetes (the Danish Council for Strategic Research, grant numbers 09-067009 and 09-075724). The study was further supported by TrygFonden, Augustinus Fonden, the Foundation for Prevention, and the Foundation for Intersectorial Research Projects. LSV was supported by a grant from the Municipality of Copenhagen, Denmark and The Danish Physiotherapist Foundation. MR-L was supported by a postdoctoral grant from the Danish Diabetes Academy supported by the Novo Nordisk Foundation and The Vissing Foundation. RC (the Parker Institute, Bispebjerg and Frederiksberg Hospital) is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL). HL was supported by a professor grant from the Municipality of Copenhagen, Denmark. The study funders were not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and the decision to submit the paper for publication.

Authors' Contributions

LSV, BKP, HL, and MR-L designed the trial and CB, JSN, KK, IKT, and AAV contributed to the trial design. IKT and LSV collected the data together with the staff at CFAS (Copenhagen, Denmark) and the participating centers and hospital. CG, JCB, and RN contributed with expert knowledge and competences. IKT, YY, RC, and MR-L planned the statistical analyses, which were conducted by IKT. IKT wrote the first draft of the manuscript. IKT and MR-L are guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statistical analysis plan.
[PDF File (Adobe PDF File), 368 KB - mhealth_v10i9e30602_app1.pdf]

Multimedia Appendix 2
Accelerometer wear time and nonwear time at baseline, 12-week, and 52-week follow-up.
[DOCX File , 17 KB - mhealth_v10i9e30602_app2.docx]

Multimedia Appendix 3
Graph of mean (SD) intensities (per minute) of all registered IWT sessions in the IWT group (ie, IWTonly + IWTsupport).
[DOCX File , 413 KB - mhealth_v10i9e30602_app3.docx]

Multimedia Appendix 4
Adherence to the intervention in the IWT combined (IWTsupport + IWTonly), IWTonly and IWTsupport groups.
[DOCX File , 17 KB - mhealth_v10i9e30602_app4.docx]

Multimedia Appendix 5
Overview of adherence to the follow-up motivational interviews for the IWT support group.
[DOCX File , 14 KB - mhealth_v10i9e30602_app5.docx]

Multimedia Appendix 6
Intention-to-treat analysis of changes from baseline to 12-week intermediate assessment between the IWT group and the StC group.
[DOCX File , 16 KB - mhealth_v10i9e30602_app6.docx]

Multimedia Appendix 7
Sensitivity analyses of changes in MVPA time from baseline to 52-week follow-up between the IWT group and the StC group: (1) Per-protocol analysis; (2) Analysis with non-responder (baseline observation carried forward; BOCF) imputation; (3) Analysis with the subsample of participants whose intervention period did not overlap time of technological malfunction of the InterWalk application; and (4) Analysis with daytime criteria for accelerometer data (ie, ≥3 days of ≥14 h of daytime measurement (6 am to 22 pm)). (5) Sensitivity analysis of changes in forced walking (ie, ≥3000 CPM) time from baseline to 52-week follow-up between the IWT group and the StC group.
[DOCX File , 14 KB - mhealth_v10i9e30602_app7.docx]

Multimedia Appendix 8
Post hoc linear regression analysis of the dose response association over 52 weeks between (a) change in MVPA (min/day) and change in SF-12 PCS score; (b) IWT duration (min/week) and change in SF-12 PCS score; (c) change in MVPA (min/day) and change in waist circumference (cm); and (d) IWT duration (min/week) and change in waist circumference (cm).
[DOCX File , 368 KB - mhealth_v10i9e30602_app8.docx]

Multimedia Appendix 9
Intention-to-treat analysis of changes from baseline to 52-week follow-up between the IWTonly group and the IWTsupport group.
[DOCX File , 17 KB - mhealth_v10i9e30602_app9.docx]

Multimedia Appendix 10
Bar chart of median (25th; 75th percentile) IWT duration (min/week) over 52 weeks among IWT subgroups of sex (men/women), type 2 diabetes duration (≤5 years/>5 years), alcohol consumption (within the recommended levels/above recommendations), smoking habits (smoker/nonsmoker), highest level of education (ISCED-2011 levels 0-4/ISCED-2011 levels 5-8), civil status (single, divorced, or widowed/married or cohabiting), SF-12 PCS (high/low), and SF-12 MCS (high/low), assessed at baseline.
[DOCX File , 121 KB - mhealth_v10i9e30602_app10.docx]

Multimedia Appendix 11
Baseline demographic and clinical characteristics among attenders and nonattenders at 52-week follow-up.
[DOCX File , 50 KB - mhealth_v10i9e30602_app11.docx]

Multimedia Appendix 12
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 9080 KB - mhealth_v10i9e30602_app12.pdf]

References

1. American Diabetes Association. 5. Facilitating Behavior Change and Well-being to Improve Health Outcomes:. Diabetes Care 2020 Jan;43(Suppl 1):S48-S65. [doi: 10.2337/dc20-S005] [Medline: 31862748]
2. Kanaley JA, Colberg SR, Corcoran MH, Malin SK, Rodriguez NR, Crespo CJ, et al. Exercise/Physical Activity in Individuals with Type 2 Diabetes: A Consensus Statement from the American College of Sports Medicine. Med Sci Sports Exerc 2022 Feb 01;54(2):353-368. [doi: 10.1249/MSS.0000000000002800] [Medline: 35029593]
3. Balducci S, Sacchetti M, Haxhi J, Orlando G, D'Errico V, Fallucca S, et al. Physical exercise as therapy for type 2 diabetes mellitus. Diabetes Metab Res Rev 2014 Mar;30 Suppl 1:13-23. [doi: 10.1002/dmrr.2514] [Medline: 24353273]
4. Booth FW, Hawley JA. The erosion of physical activity in Western societies: an economic death march. Diabetologia 2015 Aug;58(8):1730-1734. [doi: 10.1007/s00125-015-3617-5] [Medline: 25952481]
5. Beedie C, Mann S, Jimenez A, Kennedy L, Lane AM, Domone S, et al. Death by effectiveness: exercise as medicine caught in the efficacy trap!. Br J Sports Med 2016 Mar;50(6):323-324 [FREE Full text] [doi: 10.1136/bjsports-2014-094389] [Medline: 25807166]
6. Church TS, Blair SN. When will we treat physical activity as a legitimate medical therapy...even though it does not come in a pill? Br J Sports Med 2009 Feb;43(2):80-81. [doi: 10.1136/bjsm.2008.053850] [Medline: 18927165]
7. Ried-Larsen M, MacDonald CS, Johansen MY, Hansen KB, Christensen R, Almdal TP, et al. Why prescribe exercise as therapy in type 2 diabetes? We have a pill for that!. Diabetes Metab Res Rev 2018 Jul;34(5):e2999. [doi: 10.1002/dmrr.2999] [Medline: 29488311]
8. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an Intensive Lifestyle Intervention on Glycemic Control in Patients With Type 2 Diabetes: A Randomized Clinical Trial. JAMA 2017 Aug 15;318(7):637-646 [FREE Full text] [doi: 10.1001/jama.2017.10169] [Medline: 28810024]
9. Look AHEAD Research Group, Pi-Sunyer X, Blackburn G, Brancati FL, Bray GA, Bright R, et al. Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: one-year results of the look AHEAD trial. Diabetes Care 2007 Jun;30(6):1374-1383 [FREE Full text] [doi: 10.2337/dc07-0048] [Medline: 17363746]
10. Church TS, Blair SN, Cocreham S, Johannsen N, Johnson W, Kramer K, et al. Effects of aerobic and resistance training on hemoglobin A1c levels in patients with type 2 diabetes: a randomized controlled trial. JAMA 2010 Nov 24;304(20):2253-2262 [FREE Full text] [doi: 10.1001/jama.2010.1710] [Medline: 21098771]
11. Andrews RC, Cooper AR, Montgomery AA, Norcross AJ, Peters TJ, Sharp DJ, et al. Diet or diet plus physical activity versus usual care in patients with newly diagnosed type 2 diabetes: the Early ACTID randomised controlled trial. Lancet 2011 Jul 09;378(9786):129-139 [FREE Full text] [doi: 10.1016/S0140-6736(11)60442-X] [Medline: 21705068]
12. Balducci S, D'Errico V, Haxhi J, Sacchetti M, Orlando G, Cardelli P, Italian DiabetesExercise Study 2 (IDES_2) Investigators. Effect of a Behavioral Intervention Strategy on Sustained Change in Physical Activity and Sedentary Behavior in Patients

With Type 2 Diabetes: The IDES_2 Randomized Clinical Trial. JAMA 2019 Mar 05;321(9):880-890. [doi: 10.1001/jama.2019.0922] [Medline: 30835309]
13. Look AHEAD Research Group, Wing RR. Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial. Arch Intern Med 2010 Sep 27;170(17):1566-1575. [doi: 10.1001/archinternmed.2010.334] [Medline: 20876408]
14. Ried-Larsen M, Johansen MY, MacDonald CS, Hansen KB, Christensen R, Wedell-Neergaard A, et al. Type 2 diabetes remission 1 year after an intensive lifestyle intervention: A secondary analysis of a randomized clinical trial. Diabetes Obes Metab 2019 Oct;21(10):2257-2266 [FREE Full text] [doi: 10.1111/dom.13802] [Medline: 31168922]
15. Unick JL, Gaussoin SA, Hill JO, Jakicic JM, Bond DS, Hellgren M, et al. Four-Year Physical Activity Levels among Intervention Participants with Type 2 Diabetes. Med Sci Sports Exerc 2016 Dec;48(12):2437-2445 [FREE Full text] [doi: 10.1249/MSS.0000000000001054] [Medline: 27471785]
16. Umpierre D, Ribeiro PAB, Kramer CK, Leitão CB, Zucatti ATN, Azevedo MJ, et al. Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis. JAMA 2011 May 04;305(17):1790-1799. [doi: 10.1001/jama.2011.576] [Medline: 21540423]
17. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
18. Gibala MJ, Little JP. Physiological basis of brief vigorous exercise to improve health. J Physiol 2020 Jan;598(1):61-69 [FREE Full text] [doi: 10.1113/JP276849] [Medline: 31691289]
19. Karstoft K, Winding K, Knudsen SH, Nielsen JS, Thomsen C, Pedersen BK, et al. The effects of free-living interval-walking training on glycemic control, body composition, and physical fitness in type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2013 Feb;36(2):228-236 [FREE Full text] [doi: 10.2337/dc12-0658] [Medline: 23002086]
20. Ried-Larsen M, Thomsen RW, Berencsi K, Brinkløv CF, Brøns C, Valentiner LS, et al. Implementation of interval walking training in patients with type 2 diabetes in Denmark: rationale, design, and baseline characteristics. Clin Epidemiol 2016;8:201-209 [FREE Full text] [doi: 10.2147/CLEP.S97303] [Medline: 27354828]
21. Ashrafzadeh S, Hamdy O. Patient-Driven Diabetes Care of the Future in the Technology Era. Cell Metab 2019 Mar 05;29(3):564-575 [FREE Full text] [doi: 10.1016/j.cmet.2018.09.005] [Medline: 30269984]
22. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2022-04-11]
23. Valentiner LS, Ried-Larsen M, Karstoft K, Brinkløv CF, Brøns C, Nielsen RO, et al. Long-term effect of smartphone-delivered Interval Walking Training on physical activity in patients with type 2 diabetes: protocol for a parallel group single-blinded randomised controlled trial. BMJ Open 2017 Apr 07;7(4):e014036 [FREE Full text] [doi: 10.1136/bmjopen-2016-014036] [Medline: 28389489]
24. Valentiner LS, Thorsen IK, Kongstad MB, Brinkløv CF, Larsen RT, Karstoft K, et al. Effect of ecological momentary assessment, goal-setting and personalized phone-calls on adherence to interval walking training using the InterWalk application among patients with type 2 diabetes-A pilot randomized controlled trial. PLoS One 2019;14(1):e0208181 [FREE Full text] [doi: 10.1371/journal.pone.0208181] [Medline: 30629601]
25. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
26. The Capital Region of Denmark. Forløbsprogram for Type 2 Diabetes. Region Hovedstaden. URL: https://www.regionh.dk/til-fagfolk/Sundhed/Tvaersektorielt-samarbejde/kronisk-sygdom/Forl%C3%B8bsprogrammer/Sider/Forloebsprogram-for-type-2-diabetes.aspx [accessed 2022-04-11]
27. Miller W, Rollnick S. Motivational interviewing: preparing people for change. 2nd ed. New York, NY: Guilford Press; 2002.
28. Brønd JC, Aadland E, Andersen LB, Resaland GK, Andersen SA, Arvidsson D. The ActiGraph counts processing and the assessment of vigorous activity. Clin Physiol Funct Imaging 2019 Jul;39(4):276-283. [doi: 10.1111/cpf.12571] [Medline: 30980611]
29. Toftager M, Kristensen P, Oliver M, Duncan S, Christiansen L, Boyle E, et al. Accelerometer data reduction in adolescents: effects on sample retention and bias. Int J Behav Nutr Phys Act 2013 Dec 23;10(1):140 [FREE Full text] [doi: 10.1186/1479-5868-10-140] [Medline: 24359480]
30. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
31. Petersen TL, Brønd JC, Kristensen PL, Aadland E, Grøntved A, Jepsen R. Resemblance in accelerometer-assessed physical activity in families with children: the Lolland-Falster Health Study. Int J Behav Nutr Phys Act 2020 Dec 04;17(1):161 [FREE Full text] [doi: 10.1186/s12966-020-01067-7] [Medline: 33276796]
32. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]

33. Brinkløv CF, Thorsen IK, Karstoft K, Brøns C, Valentiner L, Langberg H, et al. Criterion validity and reliability of a smartphone delivered sub-maximal fitness test for people with type 2 diabetes. BMC Sports Sci Med Rehabil 2016;8:31. [doi: 10.1186/s13102-016-0056-7] [Medline: 28174664]
34. Besson H, Brage S, Jakes RW, Ekelund U, Wareham NJ. Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. Am J Clin Nutr 2010 Jan;91(1):106-114. [doi: 10.3945/ajcn.2009.28432] [Medline: 19889820]
35. Markland D, Tobin V. A Modification to the Behavioural Regulation in Exercise Questionnaire to Include an Assessment of Amotivation. Journal of Sport and Exercise Psychology 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
36. Skotte J, Korshøj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. J Phys Act Health 2014 Jan;11(1):76-84. [doi: 10.1123/jpah.2011-0347] [Medline: 23249722]
37. Ekelund U, Brage S, Griffin SJ, Wareham NJ, ProActive UK Research Group. Objectively measured moderate- and vigorous-intensity physical activity but not sedentary time predicts insulin resistance in high-risk individuals. Diabetes Care 2009 Jun;32(6):1081-1086 [FREE Full text] [doi: 10.2337/dc08-1895] [Medline: 19252168]
38. Thorsen I, Yang Y, Ried-Larsen M. Long-term Effectiveness of Smartphone App-Delivered Interval Walking Training on Physical Activity Among Individuals with Type 2 Diabetes: SAP for the parallel-group, assessor-blinded, randomized InterWalk trial. The Centre for Physical Activity Research (CFAS) Website. URL: https://aktivsundhed.dk/images/docs/SAP_Interwalk_v11.pdf [accessed 2022-04-11]
39. Cao J, Zhang S. Multiple comparison procedures. JAMA 2014 Aug 06;312(5):543-544. [doi: 10.1001/jama.2014.9440] [Medline: 25096694]
40. Díaz-Arribas MJ, Fernández-Serrano M, Royuela A, Kovacs FM, Gallego-Izquierdo T, Ramos-Sánchez M, et al. Minimal Clinically Important Difference in Quality of Life for Patients With Low Back Pain. Spine (Phila Pa 1976) 2017 Dec 15;42(24):1908-1916. [doi: 10.1097/BRS.0000000000002298] [Medline: 28658040]
41. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and Web-Based Monitoring and Feedback Tool Embedded in Primary Care Increases Physical Activity: A Cluster Randomized Controlled Trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
42. Poppe L, De Bourdeaudhuij I, Verloigne M, Shadid S, Van Cauwenberg J, Compernolle S, et al. Efficacy of a Self-Regulation-Based Electronic and Mobile Health Intervention Targeting an Active Lifestyle in Adults Having Type 2 Diabetes and in Adults Aged 50 Years or Older: Two Randomized Controlled Trials. J Med Internet Res 2019 Aug 02;21(8):e13363 [FREE Full text] [doi: 10.2196/13363] [Medline: 31376274]
43. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. A pilot study of a tool to stimulate physical activity in patients with COPD or type 2 diabetes in primary care. J Telemed Telecare 2014 Jan;20(1):29-34. [doi: 10.1177/1357633X13519057] [Medline: 24414397]
44. Plotnikoff RC, Wilczynska M, Cohen KE, Smith JJ, Lubans DR. Integrating smartphone technology, social support and the outdoor physical environment to improve fitness among adults at risk of, or diagnosed with, Type 2 Diabetes: Findings from the 'eCoFit' randomized controlled trial. Prev Med 2017 Dec;105:404-411. [doi: 10.1016/j.ypmed.2017.08.027] [Medline: 28887192]
45. MacDonald CS, Nielsen SM, Bjørner J, Johansen MY, Christensen R, Vaag A, et al. One-year intensive lifestyle intervention and improvements in health-related quality of life and mental health in persons with type 2 diabetes: a secondary analysis of the U-TURN randomized controlled trial. BMJ Open Diabetes Res Care 2021 Jan;9(1):e001840 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001840] [Medline: 33441418]
46. Ross R, Hudson R, Stotz PJ, Lam M. Effects of exercise amount and intensity on abdominal obesity and glucose tolerance in obese adults: a randomized trial. Ann Intern Med 2015 Mar 03;162(5):325-334. [doi: 10.7326/M14-1189] [Medline: 25732273]
47. Ross R, Hudson R, Day AG, Lam M. Dose-response effects of exercise on abdominal obesity and risk factors for cardiovascular disease in adults: study rationale, design and methods. Contemp Clin Trials 2013 Jan;34(1):155-160. [doi: 10.1016/j.cct.2012.10.010] [Medline: 23123790]
48. Benatti FB, Pedersen BK. Exercise as an anti-inflammatory therapy for rheumatic diseases-myokine regulation. Nat Rev Rheumatol 2015 Feb;11(2):86-97. [doi: 10.1038/nrrheum.2014.193] [Medline: 25422002]
49. Griffin SJ, Simmons RK, Prevost AT, Williams KM, Hardeman W, Sutton S, ADDITION-Plus study team. Multiple behaviour change intervention and outcomes in recently diagnosed type 2 diabetes: the ADDITION-Plus randomised controlled trial. Diabetologia 2014 Jul;57(7):1308-1319 [FREE Full text] [doi: 10.1007/s00125-014-3236-6] [Medline: 24759957]
50. Sortsø C, Lauridsen J, Emneus M, Green A, Jensen PB. Socioeconomic inequality of diabetes patients' health care utilization in Denmark. Health Econ Rev 2017 Dec;7(1):21 [FREE Full text] [doi: 10.1186/s13561-017-0155-5] [Medline: 28550486]
51. Matthews CE, Keadle SK, Troiano RP, Kahle L, Koster A, Brychta R, et al. Accelerometer-measured dose-response for physical activity, sedentary time, and mortality in US adults. Am J Clin Nutr 2016 Nov;104(5):1424-1432 [FREE Full text] [doi: 10.3945/ajcn.116.135129] [Medline: 27707702]

52. Falconer CL, Page AS, Andrews RC, Cooper AR. The Potential Impact of Displacing Sedentary Time in Adults with Type 2 Diabetes. Med Sci Sports Exerc 2015 Oct;47(10):2070-2075 [FREE Full text] [doi: 10.1249/MSS.0000000000000651] [Medline: 26378943]

Abbreviations

BREQ-2: Behavioral Regulation in Exercise Questionnaire-2
CPM: counts per minute
HRQoL: health-related quality of life
ISCED: International Standard Classification of Education
IWT: interval walking training
LPA: light-intensity physical activity
LS: least square
MCS: Mental Component Summary
MVPA: moderate-and-vigorous physical activity
PA: physical activity
PAEE: physical activity energy expenditure
PCS: Physical Component Summary
RAI: Relative Autonomy Index
RPAQ: Recent Physical Activity Questionnaire
SF-12: 12-item Short-Form Health Survey
StC: standard care
T2D: type 2 diabetees
TPA: total physical activity
VO_2peak: peak oxygen consumption

Edited by L Buis; submitted 21.05.21; peer-reviewed by D Lieberman, KL Mauco; comments to author 02.09.21; revised version received 08.02.22; accepted 22.03.22; published 28.09.22.

Please cite as:
Thorsen IK, Yang Y, Valentiner LS, Glümer C, Karstoft K, Brønd JC, Nielsen RO, Brøns C, Christensen R, Nielsen JS, Vaag AA, Pedersen BK, Langberg H, Ried-Larsen M
The Effects of a Lifestyle Intervention Supported by the InterWalk Smartphone App on Increasing Physical Activity Among Persons With Type 2 Diabetes: Parallel-Group, Randomized Trial
JMIR Mhealth Uhealth 2022;10(9):e30602
URL: https://mhealth.jmir.org/2022/9/e30602
doi:10.2196/30602
PMID:36170002

Original Paper

Using the Positive Peers Mobile App to Improve Clinical Outcomes for Young People With HIV: Prospective Observational Cohort Comparison

Mary M Step[1], MA, PhD; Jennifer McMillen Smith[2*], MSSA; Steven A Lewis[1,3*], MBA, MPH; Ann K Avery[4*], MD

[1]College of Public Health, Kent State University, Kent, OH, United States

[2]Division of Social Work, Metrohealth System, Cleveland, OH, United States

[3]Center for Health Care Research and Policy, Population Health Research Institute, Case Western Reserve University School of Medicine at The MetroHealth System, Cleveland, OH, United States

[4]Division of Infectious Diseases, Case Western Reserve University School of Medicine at The MetroHealth System, Cleveland, OH, United States

[*]these authors contributed equally

Corresponding Author:
Mary M Step, MA, PhD
College of Public Health
Kent State University
Lowry Hall, 305b
750 Hilltop Drive
Kent, OH, 44242
United States
Phone: 1 330 672 2630
Fax: 1 330 672 6505
Email: mstep@kent.edu

Abstract

Background: Disparities in HIV outcomes persist among racial, gender, and sexual minorities in the United States. Younger people face a greater risk of contracting HIV, often living without knowledge of their HIV status for long periods. The Positive Peers App (PPA) is a multifunctional HIV support tool designed to improve HIV-related clinical outcomes for young people with HIV. The app was designed according to the specifications of an in-care young adult HIV community in Northeast Ohio. Data provided in this study provide preliminary evidence of the usefulness of PPA as a relevant tool for engaging this clinical patient population in care and facilitating viral suppression.

Objective: In this study, we aimed to describe variations in PPA use and examine the associations between use and HIV clinical outcomes between self-selected user and nonuser cohorts in the same clinical population.

Methods: The PPA was offered free of charge to persons with HIV, aged 13 to 34 years of age, diagnosed with HIV within the last 12 months, out of care for 6 months during the last 24 months, or not virally suppressed (HIV viral load >200 copies/mL) in the prior 24 months. Baseline and 6- and 12-month surveys were administered via an audio computer-assisted self-interviewing system to all participants. The app's user activity was tracked natively by the app and stored in a secure server. Participant demographic and HIV care data were extracted from clinical records within 12 months before the start of the study and across the duration of the study period. HIV care outcomes of PPA users (n=114) were compared with those of nonusers (n=145) at the end of the study period (n=259).

Results: The analysis showed that younger PPA users (aged 13-24 years) were more likely to obtain HIV laboratories (adjusted odds ratio 2.85, 95% CI 1.03-7.90) and achieve sustained viral suppression than nonusers (adjusted odds ratio 4.2, 95% CI 1.2-13.9).

Conclusions: The PPA appears to help younger users sustain HIV suppression. The app offers an important tool for addressing this critical population. The PPA remains in the field and is currently being adopted by other localities to facilitate their efforts to end the HIV epidemic. Although our reported observational results require additional validation and stringent ongoing surveillance, the results represent our best efforts in a pilot study to provide a measure of efficacy for the PPA. Next steps include a large-scale evaluation of the PPA acceptability and effectiveness. Given the building evidence of user reports and outcomes, the freely available PPA could be a helpful tool for achieving Ending the HIV Epidemic goals.

(JMIR Mhealth Uhealth 2022;10(9):e37868) doi:10.2196/37868

KEYWORDS

mobile health; mHealth; HIV; Positive Peers; retention in care; re-engagement in care; viral suppression

Introduction

Background

Although clinical outcomes have improved in adults diagnosed with HIV [1], significant disparities remain for young people [2-4]. Most new HIV diagnoses in the United States are among adolescents and young adults [5]. Although ≥60% of younger (aged 13-24 years) people living with HIV are virally suppressed, significantly fewer Black and Latinx demographic groups are not [4]. Furthermore, among all young people, close to 80% are transgender or cisgender males, who most often (69%) reported HIV transmission as occurring via male-to-male sexual contact [4,5]. Given these trends, young people with HIV can experience the intersection of multiple disenfranchised communities, resulting in compounded stigma, social and family isolation, and socially determined barriers to HIV care [6-8]. The downstream effects of this burden can determine decreased lifetime health and overall longevity. Importantly, tailored mobile health interventions have been shown to effectively reduce HIV disparities for younger people and those who identify with a gender or sexual minority identity [9-12].

Mobile health apps can harness the dissemination dynamics of social media either by linking to existing platforms or by creating networks of people with similar health challenges [13]. As social media networks allow for a more user-centric, collaborative communication process, they offer greater opportunities for engagement with both health information and similar others [13,14]. However, although several studies have shown that social media platforms can serve as an effective channel for disseminating information [15,16], fewer studies link to health outcomes or identify mechanisms for change [17]. Therefore, mobile platforms that offer an effective interface for receiving tailored HIV-related information, track use, and afford users an opportunity to engage in their own recovery may have a meaningful impact on the HIV care cascade.

Prior Work

The Positive Peers App (PPA) was created as a suite of app functions that can address the range of possible needs a young person living with HIV might have [18]. Following the formation of a community advisory board, we developed specific technical features that promoted user agency to best address users' needs and provide continuous, vetted, and tailored content directly to demographically defined user groups. The resulting PPA is the center of a social media–supported network that consists of a website, Instagram, TikTok, and Twitter feed that reaches out to the HIV community with a stream of evidence-based HIV-relevant content, targeted at adolescent and young adult user groups. PPA functions range from passive to highly, including the provision of local resources (eg, housing and counseling), topical blogs, narrative accounts, medication reminders, a community forum and private chat. Evidence to date supports the PPA as being acceptable and received as intended by users [18,19]. User feedback suggests that the shared experience found among PPA users is perceived as both restorative and transitional [19]. PPA use has been linked to decreased perceived stigma, and users report that the privacy, opportunities for private instant chat, and simple self-management tools provide a useful, safe, and supportive space protected from discrimination and judgment [19].

Generally, we expect that the more a user engages with personally relevant aspects of the app, the more likely the person is to learn from the app content and form internet-based supportive relationships with other users. This prediction rests on a user-centric model of mobile app use that suggests user's needs and characteristics of the technology interact to determine user engagement [20-22]. We expect greater engagement with the mobile app to influence the acceptance of promoted HIV messaging and increase the likelihood of desirable HIV clinical outcomes [23].

Goal of This Study

Although users report liking the PPA and community [19], it is important to evaluate whether the app provides a clinical benefit to users. Our aim for this study was to determine whether PPA use provides a clinical benefit to young people living with HIV. We expect that (1) PPA users will be more engaged in care than a nonparticipating cohort from the same clinic and (2) PPA users will demonstrate greater viral suppression than those who do not use the app. In addition, we aim to learn whether relevant user demographics or personal characteristics are associated with app use or whether defined user engagement groups experience greater or lesser benefits.

Methods

Research Design

The parent study for this work was designed to develop, build, and analyze the feasibility and acceptability of the PPA by the targeted user group [18]. This study used a prospective observational single-cohort design, with measures assessed at baseline and at 6, 12, and 18 months. This study was designed after the app was introduced in the field to extend our evaluation of PPA use to HIV clinical outcomes.

App use was logged in real time and tracked directly by the operating system of the app. Clinical outcome data before and after PPA use were obtained from the electronic health record. We recognize that although a randomized controlled trial is ideal for isolating predicted effects, this pilot demonstration project was preceded by a lengthy preliminary design stage that precluded an additional clinical trial evaluation. Consequently, a cohort comparison design of PPA users and eligibility-matched nonusers within the same clinical population during the same time frame provided a reasonable option for evaluating clinical outcomes retrospectively [24] during the study period (October 2016 to May 2019).

Ethics Approval

This study was approved by the Institutional Review Board for Human Research Protections of the MetroHealth System (IRB:15-00741) on July 20, 2016, and is reviewed annually.

Participants

The study sample was derived from the HIV clinic population at a public hospital in Cleveland, Ohio, which serves as the primary source of medical care for the surrounding underserved neighborhoods in and around Cuyahoga County. Eligibility requirements for participation included (1) age between 13 and 34 years; (2) receiving HIV care within the health system; and (3) an HIV diagnosis within the last 12 months, out of care for 6 months during the last 24 months, or lack of viral suppression (HIV viral load >200 copies/mL) in the previous 24 months. Essentially, the participants were either newly diagnosed or not fully engaged in HIV care.

Study Recruitment and Comparison Cohort

Potential participants were first identified via an electronic health record query and referrals from clinic staff. At the end of recruitment, the study sample included 114 young people with HIV who remained enrolled for the duration of the study period.

A local cohort of young people with HIV not enrolled in the PPA pilot was identified for comparison with the pilot sample of PPA users. This comparison cohort (n=259) comprised patients who met the same eligibility criteria for enrollment in the parent PPA study and had a visit to the HIV clinic during the enrollment period but did not enroll to use the app or participate in the study. A manual chart review confirmed the eligibility criteria for the entire sample and provided a record of all clinic visits and laboratory results completed during the study period. It is unknown whether non-PPA user patients were invited and declined study participation or were simply not made aware of the app.

Measurement

Demographic Characteristics

Demographic information was collected from the PPA participants and comparison cohort. Variables included age, race, ethnicity, education, employment status, sexual orientation, and incarceration history, all known social determinants that influence disparities in HIV outcomes [25-29]. Age was categorized as 13 to 24, 25 to 29, and 30 to 34 years to facilitate comparisons among commonly defined age classes in HIV research [26]. Race and ethnicity were reported using the US Census Bureau categories. Finally, respondents were asked to report the number of times they were incarcerated in a jail or prison. Incarceration history was categorized for analysis as none, 1 or 2 times, or ≥3 times.

PPA Engagement

PPA engagement among app users was assessed directly from native app performance data associated with each user and stored on a secure server. Variables included the number of times the user logged in and the number of acts the user completed while logged in. Wide variability was observed across these app variables. Consequently, we categorized the number of user acts variable to better compare the user's app activity. Using the median value as a cutoff point, 3 ordinal categories of PPA use were created based on the number of actions a user took during the first 3 months of having the app on their phone: 0, none; 1, low or moderate (at or below median use); and 2, high (above median use).

HIV Outcomes

Consistent with the Health Resources Service Administration Ryan White program standards, HIV outcomes included engagement in care and HIV viral suppression [26]. Engagement in care was coded as *yes* if that had an office visit or laboratory tests completed at both 6 and 12 months for prestudy or poststudy entry. Viral suppression was coded *yes* if the viral load was less than 200 copies per ml at both 6 and 12 months following diagnosis.

Statistical Analysis

Baseline characteristics were examined across and within samples of PPA participants and the nonparticipating cohort where Pearson chi-square test was used to formally test for differences between the app user and nonuser groups. Full data were present for all characteristics except race, for which ≤3% of the values were missing.

App user activity, including the number of log-ins, features used, and number of user acts, was examined for PPA participants during the 6-month period following enrollment. Inspection of these data showed that app activity tended to diminish rapidly after 3 months. Consequently, user engagement with the app was derived from app activity in the first 3 months. Medians and IQRs were reported for the number of log-ins and the number of user acts owing to the noted skewness in distributions. Formal tests of significance were not performed because of concerns regarding the sample size.

To test the contribution of app use to outcomes, 3 separate logistic regression models were developed by regressing each HIV outcome (ie, office visits, completion of HIV laboratory tests and HIV viral suppression) on PPA participation (yes or no), while controlling for baseline characteristics and measures. Interaction effects were tested in the models to assess the potential for effect modification relative to PPA participation and individual characteristics. For each outcome modeled, odds ratios and corresponding 95% CIs were reported for PPA participation versus nonparticipation in either the overall or stratified models (in cases of significant interaction or effect modification).

In addition, outcomes were evaluated from the medical records before participation (before using the PPA) and following participation (after using the PPA) to determine if those measures had significantly changed for PPA participants. A McNemar test of agreement was used as a formal test of differences. We also examined the differences for each outcome with respect to app engagement using the categorized version of the number of user acts described earlier. Each outcome was evaluated separately for pre-PPA use and post-PPA use outcomes. Fisher exact tests were performed to test for

differences across the 3 categories of user acts. Statistical significance was determined with a *P* value cutoff of .05.

Results

Demographic Differences Between PPA User and Nonuser Comparison Groups

Demographic characteristics between enrolled PPA users and the comparison cohort at the same clinic were compared. The unenrolled group were registered patients at the host clinic who either chose not to enroll or did not learn about the study. Table 1 provides the baseline characteristics and outcomes of the study sample groups. The young people with HIV studied across the PPA user and comparison groups were predominately male (310/373, 83.1%) and Black (257/373, 70.8%). At the start of the PPA study period, 69.2% (258/373) of patients had been out of HIV care. PPA participants, relative to the comparison group, were more likely to be younger, multiracial or *other* race, and newly diagnosed.

Table 1. Baseline characteristics of Positive Peers App (PPA) participants and nonparticipant comparison groups.

Characteristic	Total (n=373), n (%)	PPA (n=114), n (%)	Non-PPA (n=259), n (%)	*P* value[a]
Age group (years)				<.001
13-24	88 (23.6)	40 (35.1)	48 (18.5)	
25-29	158 (42.4)	52 (45.6)	106 (40.9)	
30-34	127 (34.1)	22 (19.3)	105 (40.5)	
Sex at birth				.12
Male	310 (83.1)	100 (87.7)	210 (81.1)	
Female	63 (16.9)	14 (12.3)	49 (18.9)	
Race				.006
African American	257 (70.8)	78 (68.4)	179 (71.9)	
White	83 (22.9)	22 (19.3)	61 (24.5)	
Multiracial or other	23 (6.3)	14 (12.3)	9 (3.6)	
Newly diagnosed				<.001
Yes	107 (28.7)	45 (39.5)	62 (23.9)	
No, noncongenital	252 (67.6)	59 (51.8)	193 (74.5)	
No, congenital	14 (3.8)	10 (8.8)	4 (1.5)	
Out of care				.008
Yes	258 (69.2)	68 (59.7)	190 (73.4)	
No	115 (30.8)	46 (40.4)	69 (26.6)	
Office visits 6-12 months prior				.91
Yes	100 (27.2)	31 (27.2)	69 (26.6)	
No	273 (72.8)	83 (72.8)	190 (73.4)	
HIV laboratory test 6-12 months prior				.88
Yes	80 (21.5)	25 (21.9)	55 (21.2)	
No	293 (78.6)	89 (78.1)	204 (78.8)	
HIV viral suppression 6-12 months prior[b]				.27
Yes	61 (16.4)	15 (13.2)	46 (17.8)	
No	312 (83.7)	99 (86.8)	213 (82.2)	

[a]*P* values generated from Pearson χ^2 tests.

[b]HIV viral suppression defined as not detectable: <200 copies/mL.

PPA Use Across Demographic Groups

Table 2 summarizes the types of PPA used across demographic groups. A total of 81.6% (93/373) of participants logged on to the PPA during the first 3 months following enrollment. The median number of log-ins by users was 9 (IQR 4.0-18.0), and the median number of user acts was 101 (IQR 46-183). Both median number of log-ins and user acts were lowest for the oldest 30 to 34 age group (vs other age groups), and the median number of user acts was higher for White users (vs African American), single (vs in a relationship), and “nonstraight” (vs “straight”) users. Although people employed full time logged

into the app more than the other groups, unemployed participants showed the highest number of user acts. Similarly, the median number of log-ins was the highest for females, while the median number of user acts was the highest for males. Both median number of log-ins and user acts were highest for Latinx ethnicity (vs not Latinx), those newly diagnosed with HIV (vs not newly diagnosed), those carrying private or commercial insurance (vs other forms of insurance or no insurance), and those without a prior incarceration history (vs with an incarceration history).

Table 2. Positive Peers App use by patient characteristics (months 1-3).

Characteristic	Logged in? (n=114), n (%)		Number of times user logged in (n=93), median (IQR)	Number of user acts (n=93), median (IQR)
	Yes	No		
Overall	93 (81.6)	21 (18.4)	9.0 (4.0-18.0)	101.0 (46.0-183.0)
Age group (years)				
13-24	34 (85)	6 (15)	11.0 (4.0-18.0)	93.5 (48.0-161.0)
25-29	42 (80.8)	10 (19.2)	9.0 (5.0-18.0)	115.5 (65.0-183.0)
30-34	17 (77.3)	5 (22.7)	6.0 (3.0-21.0)	58.0 (30.0-214.0)
Sex at birth				
Male	82 (82)	18 (18)	8.5 (4.0-18.0)	104.0 (48.0-184.0)
Female	11 (78.6)	3 (21.4)	11.0 (2.0-19.0)	85.0 (30.0-151.0)
Race and ethnicity				
African American	61 (78.2)	17 (21.8)	8.0 (3.0-16.0)	94.0 (38.0-171.0)
White	19 (86.4)	3 (13.6)	9.0 (5.0-18.0)	140.0 (63.0-193.0)
Multiracial or other	13 (92.9)	1 (7.1)	9.0 (4.0-19.0)	88.0 (65.0-183.0)
Latinx				
Yes	12 (85.7)	2 (14.3)	12.5 (9.0-23.5)	145.5 (91.5-237.0)
No	81 (81)	19 (19)	8.0 (4.0-16.0)	93.0 (43.0-180.0)
Newly diagnosed				
Yes	38 (84.4)	7 (15.6)	12.0 (6.0-18.0)	118.5 (65.0-192.0)
No, noncongenital	46 (78)	13 (22)	6.5 (3.0-18.0)	97.5 (41.0-193.0)
No, congenital	9 (90)	1 (10)	9.0 (2.0-12.0)	65.0 (38.0-94.0)
Sexual preference				
Straight	21 (80.8)	5 (19.2)	8.0 (2.0-15.0)	70.0 (31.0-150.0)
Not straight	72 (81.8)	16 (18.2)	9.0 (5.0-18.0)	113.5 (57.0-188.0)
School completed				
HS[a] graduate	75 (83.3)	15 (16.7)	9.0 (4.0-18.0)	102.0 (45.0-184.0)
Not an HS graduate	18 (75)	6 (25)	8.0 (3.0-18.0)	93.5 (46.0-169.0)
Employment status				
Full time	26 (92.9)	2 (7.1)	11.5 (5.0-18.0)	106.0 (70.0-193.0)
Part time	21 (84)	4 (16)	5.0 (4.0-16.0)	88.0 (42.0-182.0)
Unemployed	37 (75.5)	12 (24.5)	9.0 (5.0-18.0)	113.0 (45.0-183.0)
Other[b]	9 (75)	3 (25)	4.0 (2.0-12.0)	84.0 (38.0-150.0)
Health insurers				
No insurance	10 (90.9)	1 (9.1)	8.5 (4.0-16.0)	114.5 (48.0-161.0)
Medicaid or Medicare	60 (76)	19 (24)	8.0 (4.0-15.5)	84.5 (36.0-183.5)
Private	13 (100)	0 (0)	12.0 (8.0-22.0)	171.0 (106.0-248.0)
Other insurance	9 (100)	0 (0)	9.0 (5.0-21.0)	94.0 (88.0-137.0)
Relationship status				
Single and not dating anyone	64 (83.1)	13 (16.9)	9.0 (4.0-18.5)	104.0 (44.5-199.5)
In a relationship	29 (78.4)	8 (21.6)	8.0 (3.0-16.0)	94.0 (45.0-171.0)
Incarceration				
0 times	44 (84.6)	8 (15.4)	10.5 (4.5-18.0)	104.0 (42.5-214.5)

Characteristic	Logged in? (n=114), n (%)		Number of times user logged in (n=93), median (IQR)	Number of user acts (n=93), median (IQR)
	Yes	No		
1-2 times	26 (83.9)	5 (16.1)	8.5 (5.0-14.0)	97.5 (58.0-169.0)
≥3 times	23 (74.2)	8 (25.8)	6.0 (3.0-18.0)	89.0 (31.0-310.0)

[a]HS: high school.

[b]In school, disability.

HIV Outcomes Between PPA Users and Nonusers

We were interested in determining whether there were significant differences across HIV outcomes within our cohort of PPA users. Table 3 presents the results of regressing 6- to 12-month postbaseline office visit attendance on PPA participation while adjusting for baseline characteristics and prebaseline medical records. The data suggest that no significant differences in clinic attendance exist based on the PPA participation status. In other words, the clinical comparison group was no different than the group of app users in prestudy period clinic attendance. After the app use period, there were no significant differences in PPA participation in HIV clinical outcomes based on race or new diagnosis or out-of-care status. However, among demographic variables, an interaction effect was detected for age. Age-stratified results suggested significantly improved outcomes for PPA users in the youngest (13-24 years) age group. Across the youngest age groups, PPA participants were more likely to obtain their HIV laboratory tests (adjusted odds ratio 2.85, 95% CI 1.03-7.90) than same-age nonparticipants. Importantly, younger PPA participants were also more likely to be virally suppressed (adjusted odds ratio 4.22, 95% CI 1.28-13.89) compared with nonparticipants. No significant differences in PPA participation in either HIV laboratory tests or viral suppression were observed among the older age groups.

Table 3. HIV outcomes between Positive Peers App (PPA) users and non-PPA comparison cohort.

	Outcomes[a], aOR[b] (95% CI)		
	Office visits	HIV laboratory tests	HIV viral suppression
All patients[c]			
PPA	1.66 (0.99-2.80)	—[d]	—
Non-PPA	Reference	—	—
Age 13-24 years[e]			
PPA	—	2.85 (1.03-7.90)	4.22 (1.28-13.89)
Non-PPA	—	Reference	Reference
Age 25-29 years			
PPA	—	1.86 (0.84-4.12)	1.07 (0.46-2.50)
Non-PPA	—	Reference	Reference
Age 30-34 years			
PPA	—	0.54 (0.18-1.64)	0.45 (0.11-1.75)
Non-PPA	—	Reference	Reference

[a]Outcomes are measured 6-12 months after baseline measure.

[b]aOR: adjusted odds ratio.

[c]Overall model adjusted for age, sex at birth, race, newly diagnosed, out of care, and 6-12 months prebaseline measures.

[d]When modeling outcomes, all-patient models are not relevant when age acts as an effect modifier, and age-stratified models are not relevant when age does not act as an effect modifier.

[e]Age-stratified models adjusted for sex at birth, race, newly diagnosed, out of care, and 6-12 months prebaseline measures.

Effects of PPA Use on HIV Outcomes

Tables 4 and 5 illustrate that engagement in care and HIV viral suppression significantly improved following participation in the PPA (27.2% vs 52.6%, 21.9% vs 45.6%, and 13.2% vs 29.8%). Interestingly, before PPA enrollment, eventual high users of the app were less likely to have had office visits (7/43, 15%), to have a HIV laboratory test drawn (4/43, 9%), or to have been virally suppressed (2/43, 4%) than eventual nonusers or low to moderate users. As shown in the prestudy columns, before downloading the PPA, only 15% (7/43) of high users had office visits compared with 33.3% (7/114) and 36% (17/43) for eventual nonusers and low to moderate users, respectively. However, following PPA participation, these groups converged (11/21, 52%; 26/47, 55.3; vs 23/46, 50%), suggesting that app

use facilitated engagement in care and viral suppression for patients most out of compliance.

Table 4. HIV outcomes (Office visits and HIV labs) before and after Positive Peers App (PPA) participation and stratified by app use (N=114).

Office visits?	Pre-PPA use	Post-PPA use	*P* value[a]	HIV laboratory tests?	Pre-PPA use	Post-PPA use	*P* value
Overall, n (%)			<.001				<.001
Yes	31 (27.2)	60 (52.6)		Yes	25 (21.9)	52 (45.6)	
No	83 (72.8)	54 (47.4)		No	89 (78.1)	62 (54.4)	
Number of user acts, n (%)							
None (n=21)							
Yes	7 (33.3)	11 (52.4)	—[b]	Yes	3 (14.3)	9 (42.9)	—
No	14 (66.7)	10 (47.6)	—	No	18 (85.7)	12 (57.1)	—
Low-moderate (n=47)							
Yes	17 (36.2)	26 (55.3)	—	Yes	18 (38.3)	21 (44.7)	—
No	30 (63.8)	21 (44.7)	—	No	29 (61.7)	26 (55.3)	—
High (n=46)							
Yes	7 (15.2)	23 (50.0)	—	Yes	4 (8.7)	22 (47.8)	—
No	39 (84.8)	23 (50.0)	—	No	42 (91.3)	24 (52.2)	—
P value[c]	.05	.91	—	—	.002	.91	—

[a]*P* values compare HIV outcomes before and after PPA participation using the McNemar test of agreement.

[b]Not applicable.

[c]*P* values compare the categories of user activity against Pre-PPA Post-PPA HIV outcomes using the Fisher exact test.

Table 5. HIV Viral suppression before and after Positive Peers App (PPA) participation and stratified by app use (N=114).

Viral suppression?	Pre-PPA use	Post-PPA use	*P* value[a]
Overall, n (%)			<.001
Yes	15 (13.2)	34 (29.8)	—[b]
No	99 (86.8)	80 (70.2)	—
Number of user acts, n (%)			
None (n=21)			
Yes	2 (9.5)	5 (23.8)	—
No	19 (90.5)	16 (76.2)	—
Low to moderate (n=47)			
Yes	11 (23.4)	16 (34.0)	—
No	36 (76.6)	31 (66.0)	—
High (n=46)			
Yes	2 (4.4)	13 (28.3)	—
No	44 (95.6)	13 (71.7)	—
P value[c]	—	.02	.72	—

[a]*P* values compare HIV outcomes before and after PPA participation using the McNemar test of agreement.

[b]Not applicable.

[c]*P* values compare the categories of user activity against Pre-PPA Post-PPA HIV outcomes using the Fisher exact test.

Discussion

Principal Findings

This study aimed to assess the effects of PPA use on HIV clinical outcomes within a sample of out-of-care or newly diagnosed young people in one clinic community. Following at least 6 months of intervention participation, the youngest patient group (age 13-24 years) was more likely than the same-age nonuser cohort to see an improvement in completing laboratory tests and achieving or sustaining viral suppression. This is a significant finding because younger people are not only more likely to be unsuppressed and out of care but also burdened by psychosocial and behavioral risks that define this group as a high-impact target population for ending the HIV/AIDS epidemic [30,31].

It is notable that most PPA use occurred within the first 3 months of downloading the app. Although novelty may draw users to download and try an app, engagement quickly peaks as a function of continued exposure [32,33]. It may be that the knowledge gained from app engagement bolsters greater user health self-management. Specialized health apps such as PPA may facilitate positive habits such as tracking medication or marshaling support during difficult times. However, mobile app designers may also need to address novelty effects on study retention and design their recruitment and prospective data collection accordingly.

Demographic Differences in Use

The most frequent PPA users were aged <30 years, newly diagnosed, and White or Latinx. Females logged into the app more often, but males engaged in more acts overall while using the app, possibly suggesting more surfing within the app or less purposeful use [34]. For example, females may prefer to use the app primarily to meet relational needs by using private chat or community forum functions. We caution that these demographic differences may be unique to a single-site location and should not be used for generalization to larger groups. However, these findings provide a reason to further explore these patterns across sites with a larger sample size. They could be markers of more complex behavioral patterns associated with mobile app uses and effects.

Among all PPA users, engagement in care and viral suppression improved 6 months from baseline, particularly for those in the highest app use category. Although we supported our prediction that greater involvement with the app would lead to greater effects, the mechanism underlying these effects remains obscured. Mere exposure to information is insufficient for explaining complex behavioral processes and outcomes [35]. Emerging models of user engagement with interactive digital media suggest that physical interaction with, and positive perceptions of, the technology interface predict greater cognitive involvement with provided content, which in turn motivates a user's intent to manage, apply, and share that content [20,23,36]. Using this theoretical framework, we plan to conduct future research that will allow us to identify distinct use patterns over time to better understand how the PPA can be enhanced to support HIV self-management. We believe that freely available mobile apps such as the PPA could serve as a significant tool for managing the changing health and support needs as young people living with HIV adjust to a postdiagnosis life.

Limitations

There are limitations associated with this work. Research volunteers may have been more motivated to maintain adherence, regardless of their enrollment in the mobile app. The app was assertively promoted in the clinic to all potential volunteers for 2 years. There are many reasons a patient may choose not to download a health app, including privacy concerns, data download costs, and past experience with apps. We had no quantitative data on people's reasons for participating or declining participation. However, selection bias is a known risk in observational studies and may be evident if PPA users and nonusers were different in notable ways. Data from Table 1 suggest that nonusers skewed older and were more likely to be returning after a lapse in care rather than as a newly diagnosed patient. People returning to HIV care are experiencing a different affective or psychosocial experience than newly diagnosed younger adults. We need to determine whether these differences are sufficient to preclude these patients from considering using the PPA. In addition, it may be that young adults aged >25 years are more engaged in employment, established relationships, or other life responsibilities, reducing free time for app use. Finally, there were significant and unexpected differences in app use across races. Additional qualitative or mixed methods studies may reveal the nature of these differences and allow us to craft messaging and in-app tools tailored to their needs. Future analysis beyond this single clinic population will allow us to better determine additional interaction effects that may be contributing to this pattern.

Finally, a limitation is also found in the small pilot study sample size, limiting our ability to fully adjust data outcomes to additional relevant influences. The results are representative of our local public hospital community only. However, we are currently in the field collecting a larger and more diverse sample that will enable greater precision in our estimates. Along these lines, a randomized controlled trial design will present a more robust test of app impact. A key concern for future research is the determination of a potential threshold for app engagement to facilitate positive outcomes. Nevertheless, the results reported here confirm the usefulness of the PPA as a supportive tool for young people living with HIV. PPA engagement may be occurring at a particularly formative time following a new diagnosis or return to responsible care. In this way, the PPA is a feasible and effective patient-centered tool for facilitating engagement in care at a crucial turning point in HIV treatment.

Implications

This study is predicated on the idea that greater engagement with a mobile app will facilitate the likelihood of achieving desirable HIV clinical outcomes. This idea of engagement or involvement in content is fundamental to modeling communication processes and effects [37]. Our findings support this hypothesis. Although the relationship between targeted digital content engagement and positive outcomes was supported, how that process occurs remains unseen. The cognitive or affective processes inherent in digital message

processing are at the core of this process but not the sole determinant of a given outcome.

Current explanations of media use and effects are inclusive of the characteristics and features of various technologies to explain and ultimately predict outcomes [17,21,23]. Different technologies afford different experiences to users. Unlike general social media apps, the PPA affords users opportunities to tailor their anonymity as their confidence in the community grows, provides frequent social and medical information vetted by credible clinicians, and offers a supportive and monitored community that shares similar life experiences. Although there is variability across app users in terms of support needs and adjustment to living with HIV, these user differences can be intentionally targeted in app messaging and content [38,39]. This model frames our ongoing work and holds implications for future theorizing of mobile app uses and effects.

This study also has implications for HIV care. Table 1 shows the nonuser comparison cohort to include significantly more people returning to care after a lapse in the medical management of HIV. This returning-to-care population may face different psychosocial challenges as they resume an adherent lifestyle. Continued targeting of this group may address a noted literature gap regarding best practices for re-engagement in care [40-42].

Importantly, the youngest set of PPA users was most likely to realize positive outcomes 3 months after enrollment. This population of people with HIV is of greatest concern to ending the epidemic efforts. Data presented here suggest that the app had a positive impact on these groups. Counselors and clinics who care for people living with HIV need socially relevant tools for younger patients. This is particularly relevant for rural communities or young people with HIV who desire remote, around-the-clock community support. Within these communities, the PPA offers acceptance, tangible support, self-management tools, and credible HIV-relevant information in one place.

Conclusions

HIV self-management is a significant challenge for young people that can be alleviated with the use of mobile apps that bring health information, tools, and supports directly to wherever they may be. Given that new HIV cases are predominantly among younger people, this approach is crucial for achieving an undetectable HIV status for young people living with HIV. Furthermore, acceptance into a knowing and supportive community may be a key resource for increasing HIV literacy and lessening internalized stigma [42]. The PPA is currently available via Google and iOS app stores, although users are required to verify their age and diagnosis via an electronic onboarding system. These data, taken with previously published results, point to the PPA as a useful tool for helping young people living with HIV achieve clinical outcomes that will both preserve their health and contribute to ending the epidemic.

Acknowledgments

The authors would like to thank the Positive Peers App community for their willingness, enthusiasm, and relentless optimism. The authors would also like to acknowledge the support and collaboration of many professionals of the MetroHealth System who helped make this work possible. This research was funded by the cooperative agreement number H97HA28892 from the US Department of Health and Human Services, Health Resources and Services Administration, and HIV/AIDS Special Projects of National Significance program. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the government.

Conflicts of Interest

None declared.

References

1. Moore RD, Keruly JC, Bartlett JG. Improvement in the health of HIV-infected persons in care: reducing disparities. Clin Infect Dis 2012 Nov;55(9):1242-1251 [FREE Full text] [doi: 10.1093/cid/cis654] [Medline: 23019271]
2. Allan-Blitz L, Mena LA, Mayer KH. The ongoing HIV epidemic in American youth: challenges and opportunities. Mhealth 2021;7:33 [FREE Full text] [doi: 10.21037/mhealth-20-42] [Medline: 33898602]
3. Crepaz N, Hess KL, Purcell DW, Hall HI. Estimating national rates of HIV infection among MSM, persons who inject drugs, and heterosexuals in the United States. AIDS 2019 Mar 15;33(4):701-708. [doi: 10.1097/QAD.0000000000002111] [Medline: 30585840]
4. HIV surveillance supplemental report. Centers for Disease Control and Prevention. 2020. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2021-01-05]
5. HIV in the United States by Age. Centers for Disease Control and Prevention. 2021 May. URL: https://www.cdc.gov/hiv/group/age/youth/index.html [accessed 2021-09-22]
6. America's HIV Epidemic Dashboard (AHEAD). AHEAD. URL: https://ahead.hiv.gov/data/viral-suppression [accessed 2021-09-08]
7. Reisner SL, Jadwin-Cakmak L, White Hughto JM, Martinez M, Salomon L, Harper GW. Characterizing the HIV prevention and care continua in a sample of transgender youth in the U.S. AIDS Behav 2017 Dec;21(12):3312-3327 [FREE Full text] [doi: 10.1007/s10461-017-1938-8] [Medline: 29138982]

8. Shacham E, Estlund AL, Tanner AE, Presti R. Challenges to HIV management among youth engaged in HIV care. AIDS Care 2017 Feb;29(2):189-196. [doi: 10.1080/09540121.2016.1204422] [Medline: 27397139]
9. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Rossi A, Carey MP. Development and preliminary pilot testing of a peer support text messaging intervention for HIV-infected Black men who have sex with men. J Acquir Immune Defic Syndr 2017 Feb 01;74 Suppl 2:S121-S127 [FREE Full text] [doi: 10.1097/QAI.0000000000001241] [Medline: 28079722]
10. Babel RA, Wang P, Alessi EJ, Raymond HF, Wei C. Stigma, HIV risk, and access to HIV prevention and treatment services among men who have sex with men (MSM) in the united states: a scoping review. AIDS Behav 2021 Nov;25(11):3574-3604 [FREE Full text] [doi: 10.1007/s10461-021-03262-4] [Medline: 33866444]
11. Arayasirikul S, Turner C, Trujillo D, Le V, Wilson EC. Efficacy and impact of digital HIV care navigation in young people living with HIV in San Francisco, California: prospective study. JMIR Mhealth Uhealth 2020 May 08;8(5):e18597 [FREE Full text] [doi: 10.2196/18597] [Medline: 32383680]
12. Tanner AE, Song EY, Mann-Jackson L, Alonzo J, Schafer K, Ware S, et al. Preliminary impact of the weCare social media intervention to support health for young men who have sex with men and transgender women with HIV. AIDS Patient Care STDS 2018 Nov;32(11):450-458 [FREE Full text] [doi: 10.1089/apc.2018.0060] [Medline: 30398955]
13. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
14. Craig E, Wright KB. Computer-mediated relational development and maintenance on Facebook®. Commun Res Report 2012 Apr;29(2):119-129. [doi: 10.1080/08824096.2012.667777]
15. Walther JB, Jang J. Communication processes in participatory websites. J Comput Mediat Comm 2012 Oct 10;18(1):2-15. [doi: 10.1111/j.1083-6101.2012.01592.x]
16. Susarla A, Oh J, Tan Y. Social networks and the diffusion of user-generated content: evidence from YouTube. Inform Syst Res 2012 Mar;23(1):23-41. [doi: 10.1287/isre.1100.0339]
17. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. J Biomed Inform 2013 Dec;46(6):957-969 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.010] [Medline: 23702104]
18. Step MM, McMillen Smith J, Kratz J, Briggs J, Avery A. "Positive Peers": function and content development of a mobile app for engaging and retaining young adults in HIV care. JMIR Form Res 2020 Jan 30;4(1):e13495 [FREE Full text] [doi: 10.2196/13495] [Medline: 32012035]
19. Levy MR, Windahl S. Audience activity and gratifications. Commun Res 2016 Jun 29;11(1):51-78. [doi: 10.1177/009365084011001003]
20. Sundar SS, Limperos AM. Uses and grats 2.0: new gratifications for new media. J Broadcasting Electron Media 2013 Oct;57(4):504-525. [doi: 10.1080/08838151.2013.845827]
21. Phua J, Jin SV, Kim J. Uses and gratifications of social networking sites for bridging and bonding social capital: a comparison of Facebook, Twitter, Instagram, and Snapchat. Comput Human Behav 2017 Jul;72:115-122. [doi: 10.1016/j.chb.2017.02.041]
22. Oh J, Bellur S, Sundar SS. Clicking, assessing, immersing, and sharing: an empirical model of user engagement with interactive media. Commun Res 2015 Sep 21;45(5):737-763. [doi: 10.1177/0093650215600493]
23. Step MM, Knight K, McMillen Smith J, Lewis SA, Russell TJ, Avery AK. Positive Peers mobile application reduces stigma perception among young people living with HIV. Health Promot Pract 2020 Sep;21(5):744-754. [doi: 10.1177/1524839920936244] [Medline: 32757838]
24. Monitoring and Evaluating Digital Health Interventions A Practical Guide to Conducting Research and Assessment. Geneva: World Health Organization; 2016.
25. Johnson Lyons S, Gant Z, Jin C, Dailey A, Nwangwu-Ike N, Satcher Johnson A. A census tract-level examination of differences in social determinants of health among people with HIV, by race/ethnicity and geography, United States and Puerto Rico, 2017. Public Health Rep 2022;137(2):278-290. [doi: 10.1177/0033354921990373] [Medline: 33629905]
26. HIV/AIDS bureau performance measures. HRSA. 2019. URL: https://ryanwhite.hrsa.gov/sites/default/files/ryanwhite/grants/core-measures.pdf [accessed 2022-09-12]
27. Rodrigues A, Struchiner CJ, Coelho LE, Veloso VG, Grinsztejn B, Luz PM. Late initiation of antiretroviral therapy: inequalities by educational level despite universal access to care and treatment. BMC Public Health 2021 Feb 19;21(1):389 [FREE Full text] [doi: 10.1186/s12889-021-10421-8] [Medline: 33607975]
28. Maulsby CH, Ratnayake A, Hesson D, Mugavero MJ, Latkin CA. A scoping review of employment and HIV. AIDS Behav 2020 Oct;24(10):2942-2955 [FREE Full text] [doi: 10.1007/s10461-020-02845-x] [Medline: 32246357]
29. Khan MR, McGinnis KA, Grov C, Scheidell JD, Hawks L, Edelman EJ, et al. Past year and prior incarceration and HIV transmission risk among HIV-positive men who have sex with men in the US. AIDS Care 2019 Mar;31(3):349-356 [FREE Full text] [doi: 10.1080/09540121.2018.1499861] [Medline: 30064277]
30. Laurenzi CA, Toit du S, Ameyan W, Melendez-Torres GJ, Kara T, Brand A, et al. Psychosocial interventions for improving engagement in care and health and behavioural outcomes for adolescents and young people living with HIV: a systematic review and meta-analysis. J Int AIDS Soc 2021 Aug 02;24(8):e25741 [FREE Full text] [doi: 10.1002/jia2.25741] [Medline: 34338417]

31. Wood SM, Dowshen N, Lowenthal E. Time to improve the global human immunodeficiency virus/AIDS care continuum for adolescents: a generation at stake. JAMA Pediatr 2015 Jul 01;169(7):619-620 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.58] [Medline: 25985061]
32. Berlyne D. Novelty and curiosity as determinants of exploratory behaviour. Br J Psychol 1950;41(1):68-80.
33. Dubey R, Griffiths TL. Reconciling novelty and complexity through a rational analysis of curiosity. Psychol Rev 2020 Apr;127(3):455-476. [doi: 10.1037/rev0000175] [Medline: 31868394]
34. Maddux WW, Brewer MB. Gender differences in the relational and collective bases for trust. Group Processes Intergroup Relation 2005 Apr 12;8(2):159-171. [doi: 10.1177/1368430205051065]
35. Montoya RM, Horton RS, Vevea JL, Citkowicz M, Lauber EA. A re-examination of the mere exposure effect: the influence of repeated exposure on recognition, familiarity, and liking. Psychol Bull 2017 May;143(5):459-498. [doi: 10.1037/bul0000085] [Medline: 28263645]
36. Coyne SM, Padilla-Walker LM, Howard E. Emerging in a digital world. Emerg Adulthood 2013 Mar 26;1(2):125-137. [doi: 10.1177/2167696813479782]
37. Valkenburg, PM. Theoretical foundations of social media uses and effects. In: Nesi J, Teller EH, Prinstein MJ, editors. Handbook of Adolescent Digital Media Use and Mental Health. Cambridge, United Kingdom: Cambridge University Press; Jun 2022:39-60.
38. Volkom M, Stapley J, Amaturo V. Revisiting the digital divide: generational differences in technology use in everyday life. N Am J Psychol 2014;16(3):557-574 [FREE Full text]
39. Higa DH, Crepaz N, Mullins MM, Prevention Research Synthesis Project. Identifying best practices for increasing linkage to, retention, and re-engagement in HIV medical care: findings from a systematic review, 1996-2014. AIDS Behav 2016 May;20(5):951-966 [FREE Full text] [doi: 10.1007/s10461-015-1204-x] [Medline: 26404014]
40. Henkhaus ME, Hussen SA, Brown DN, Del Rio C, Fletcher MR, Jones MD, et al. Barriers and facilitators to use of a mobile HIV care model to re-engage and retain out-of-care people living with HIV in Atlanta, Georgia. PLoS One 2021;16(3):e0247328 [FREE Full text] [doi: 10.1371/journal.pone.0247328] [Medline: 33705421]
41. Risher KA, Kapoor S, Daramola AM, Paz-Bailey G, Skarbinski J, Doyle K, et al. Challenges in the evaluation of interventions to improve engagement along the HIV care continuum in the united states: a systematic review. AIDS Behav 2017 Jul;21(7):2101-2123 [FREE Full text] [doi: 10.1007/s10461-017-1687-8] [Medline: 28120257]
42. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]

Abbreviations

PPA: Positive Peers App

Edited by L Buis; submitted 09.03.22; peer-reviewed by J Jones, K Schafer; comments to author 02.05.22; revised version received 01.08.22; accepted 03.08.22; published 28.09.22.

Please cite as:
Step MM, McMillen Smith J, Lewis SA, Avery AK
Using the Positive Peers Mobile App to Improve Clinical Outcomes for Young People With HIV: Prospective Observational Cohort Comparison
JMIR Mhealth Uhealth 2022;10(9):e37868
URL: https://mhealth.jmir.org/2022/9/e37868
doi:10.2196/37868
PMID:36170001

Original Paper

Testing the Pragmatic Effectiveness of a Consumer-Based Mindfulness Mobile App in the Workplace: Randomized Controlled Trial

Jennifer L Huberty[1,2], PhD; Hallie M Espel-Huynh[1], PhD; Taylor L Neher[3], MPH; Megan E Puzia[4], MS

[1]Calm.com, Inc., San Francisco, CA, United States

[2]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[3]Center for the Study of Aging, Boise State University, Boise, ID, United States

[4]Behavioral Research and Analytics, LLC, Salt Lake City, UT, United States

Corresponding Author:
Jennifer L Huberty, PhD
Calm.com, Inc.
77 Geary St. Fl. 3
San Francisco, CA, 94108
United States
Phone: 1 402 301 1304
Email: jenhubertyphd@gmail.com

Abstract

Background: Mental health and sleep problems are prevalent in the workforce, corresponding to costly impairment in productivity and increased health care use. Digital mindfulness interventions are efficacious in improving sleep and mental health in the workplace; however, evidence supporting their pragmatic utility, potential for improving productivity, and ability to reduce employer costs is limited.

Objective: This pragmatic, cluster randomized controlled trial aimed to evaluate the experimental effects of implementing a commercially available mindfulness app—Calm—in employees of a large, multisite employer in the United States. Outcomes included mental health (depression, anxiety, and stress), sleep (insomnia and daytime sleepiness), resilience, productivity impairment (absenteeism, presenteeism, overall work impairment, and non–work activity impairment), and health care use (medical visit frequency).

Methods: Employees were randomized at the work site to receive either the Calm app intervention or waitlist control. Participants in the Calm intervention group were instructed to use the Calm app for 10 minutes per day for 8 weeks; individuals with elevated baseline insomnia symptoms could opt-in to 6 weeks of sleep coaching. All outcomes were assessed every 2 weeks, with the exception of medical visits (weeks 4 and 8 only). Effects of the Calm intervention on outcomes were evaluated via mixed effects modeling, controlling for relevant baseline characteristics, with fixed effects of the intervention on outcomes assessed at weeks 2, 4, 6, and 8. Models were analyzed via complete-case and intent-to-treat analyses.

Results: A total of 1029 employees enrolled (n=585 in the Calm intervention group, including 101 who opted-in to sleep coaching, and n=444 in waitlist control). Of them, 192 (n=88 for the Calm intervention group and n=104 for waitlist) completed all 5 assessments. In the complete-case analysis at week 8, employees at sites randomized to the Calm intervention group experienced significant improvements in depression (P=.02), anxiety (P=.01), stress (P<.001), insomnia (P<.001), sleepiness (P<.001), resilience (P=.02), presenteeism (P=.01), overall work impairment (P=.004), and nonwork impairment (P<.001), and reduced medical care visit frequency (P<.001) and productivity impairment costs (P=.01), relative to the waitlist control. In the intent-to-treat analysis at week 8, significant benefits of the intervention were observed for depression (P=.046), anxiety (P=.01), insomnia (P<.001), sleepiness (P<.001), nonwork impairment (P=.04), and medical visit frequency (P<.001).

Conclusions: The results suggest that the Calm app is an effective workplace intervention for improving mental health, sleep, resilience, and productivity and for reducing medical visits and costs owing to work impairment. Future studies should identify optimal implementation strategies that maximize employee uptake and large-scale implementation success across diverse, geographically dispersed employers.

Trial Registration: ClinicalTrials.gov NCT05120310; https://clinicaltrials.gov/ct2/show/NCT05120310

(JMIR Mhealth Uhealth 2022;10(9):e38903) doi:10.2196/38903

KEYWORDS

mindfulness; mobile apps; workforce; workplace; presenteeism; mental health

Introduction

Mental Health Concerns in the Workplace

Approximately 8% to 15% of the US adult population is affected by depression and anxiety, respectively [1,2], at least 1 in 4 is affected negatively by stress, and ≥60% report sleep disturbance such as trouble falling or staying asleep, insufficient sleep duration, or restlessness during the night [3]. Poor mental health (ie, depression, anxiety, or stress) and lack of sleep are linked to limitations in daily activities and increased health care use and thus pose a major burden to society [4-13]. Poor mental health and sleep disturbance are increasingly recognized as top contributors to reduced productivity and increased economic burden for employers (ie, via increased medical costs per employee and financial losses due to reduced work productivity) [14-16]. In the United States, depressive symptoms alone account for an estimated US $6 billion total in missed work costs (absenteeism) and nearly US $85 billion in productivity losses due to on-the-job impairment (presenteeism) [16,17]. Sleep-related productivity losses are estimated to cost more than US $1967 *per employee* each year [16].

Benefits of Mindfulness and Gaps in the Literature

Mindfulness-based workplace interventions (eg, mindfulness meditation or mindfulness-based stress reduction programs) have been shown to improve mental health (ie, depression, anxiety, and stress) and sleep [18-20]. Two recent meta-analyses examined the effects of mindfulness-based interventions in the workplace on a number of employee mental health outcomes and sleep. Mindfulness interventions have demonstrated significant benefits in anxiety, stress, and sleep [18,20]. One of the meta-analyses reported marginal improvements in depressive symptoms as a result of mindfulness interventions but noted limitations such as the small number of studies available and data inadequacy [20]. The other reported on another key construct in emotional coping–resilience–which refers to one's ability to *bounce back* or recover from stress [21,22] and which can serve as a buffer between stressful life events and the development of mental health problems including depression, anxiety, and chronic stress [23]. Meta-analytic results indicated significant effects of mindfulness interventions on resilience in the workplace, although the results were interpreted with caution because only 4 studies contributed to the analysis [22].

Even fewer studies have examined the effects on key work productivity measures, such as absenteeism (productivity losses owing to missed time working as a result of mental or physical health concerns), presenteeism (productivity losses due to impairment while on the job as a result of mental or physical health concerns), overall impairment in work productivity (absenteeism and presenteeism combined), and overall impairment in nonwork activities [18,22,24]; all of these constructs are particularly relevant to balancing the costs and benefits of these programs for employers. Some studies have found positive effects of mindfulness training on productivity [25,26], particularly for presenteeism, whereas others have found no significant impact [27,28]. One potential reason for these mixed results is the heterogeneity of measurement and the use of assessment measures that have not been extensively validated [22]. There is also a dearth of evidence for the ultimate cost-effectiveness of these programs in terms of health care expenditures for employers [18,22]. Thus, more work is needed in this area to understand the economic benefits of mindfulness-based interventions in the workplace, particularly with well-validated measures of overall productivity.

In addition, to date, most mindfulness-based interventions in the workplace have been delivered in-person, which limits scalability because of the costly need for a trained interventionist at each session [29], as well as employee-level barriers such as the need for child care, transportation to a site, and limited scheduling options [30,31]. Digital mindfulness-based interventions in the workplace offer employers the ability to help employees manage their mental health, while mitigating many barriers associated with in-person interventions. Although few studies have evaluated mindfulness apps specifically in the workplace, preliminary evidence supports their effectiveness both for mental health and productivity outcomes. For example, 1 recent study found that employees randomized to receive a commercially available mindfulness meditation app in the workplace had significantly greater improvements in anxiety and depression versus those in a waitlist control condition, but no effects on sleep were evaluated [32]. The study also found positive effects on self-reported job strain but did not assess employee productivity [32]. Another limitation is that the study sample was restricted to a relatively healthy population due to the exclusion of employees with several common physical and mental health concerns (eg, depression, hypertension, and cardiovascular disease) [32]. It is unclear how the results would generalize to the broader population of working individuals with a range of health concerns. Another recent study examined the effects of a mindfulness app in the workplace compared with an app-plus-group intervention or waitlist control; however, this study did not evaluate mental health or sleep outcomes aside from post hoc participant perceptions [28]. In addition, there was only a marginal effect on measured overall productivity, and the study did not examine presenteeism and absenteeism separately [28]. To date, no studies have assessed the effects of a stand-alone mindfulness app concurrently on mental health outcomes, resilience, workplace productivity, and health care use, particularly with an inclusive sample that is more representative of the present-day workforce.

Study Aims

The purpose of this study was to evaluate the experimental effects of the Calm app, a mobile meditation app, on employee mental health (depression, anxiety, and stress), sleep, resilience, work productivity outcomes (absenteeism, presenteeism, overall work impairment, and overall activity impairment) and health

care use (operationalized as the number of visits with a medical provider) in a workplace setting. We also explored the financial benefit of the Calm app in terms of productivity impairment cost savings (the amount of money saved per employee owing to the improvements in overall work productivity observed in the Calm intervention group vs waitlist control). This is the first study to evaluate the Calm app in a workplace setting (offered to employees for free and paid by the employer), specifically with a focus on employee mental health and productivity.

Methods

Ethics Approval

This study was approved by the institutional review board of Arizona State University (STUDY00014072) and registered with ClinicalTrials.gov (trial registration NCT05120310). All participants provided electronic informed consent before participating in the study.

Participants and Recruitment

Participants were employees of a large consumer electronics retailer. Recruitment occurred nationally between August and December 2021 via email invitations from human resources, store leaders, and flyers posted in store breakrooms. Email materials and flyers included a QR code and website link that directed participants to a web-based eligibility survey (via the Qualtrics web-based survey platform). All recruitment materials referred to the intervention as the use of a *health and wellness app*. Invitations and flyers for recruitment were sent to 294 (estimated 20,000 employees), 288 (estimated 18,000 employees), and 511 work sites (estimated 36,000 employees) over the 6-month recruitment period.

Employees were eligible for the study if they were (1) were a current employee of the company, (2) were at least 18 years of age, (3) were able to read and understand English, (4) owned a smartphone, (5) were willing to download the Calm app, and (6) were meditation-naive and had not practiced meditation for ≥60 minutes per month for the past 6 months. We included only those that were meditation-naive because the literature suggests differences in the effects of mindfulness in those that have less meditation experience than those that have more (eg, recruiting different brain regions during meditation, which are differentially related to building attentional control and emotional regulation, versus maintaining existing networks [33]). Eligibility surveys took approximately 2 minutes to complete. At the end of the survey, ineligible employees were notified of their status, and eligible employees were automatically directed to a link containing the electronic informed consent and a video explaining the details of the study and the consent form. After consenting, participants were directed to complete the baseline questionnaires.

Depending on site randomization, employees were assigned to either the intervention group (ie, Calm app [10 minutes per day]) or the waitlist control group (ie, received access to the Calm app after 8 weeks). The primary and secondary outcomes of this study were sleep (ie, insomnia symptoms, daytime sleepiness, and sleep diaries), productivity (ie, absenteeism, presenteeism, work impairment, and activity impairment), resilience, and mental health (ie, depression, anxiety, and stress). Study outcomes were assessed at baseline (week 0), midintervention (weeks 2, 4, and 6), and after the intervention (week 8). A subsample of participants who self-reported elevated sleep disturbance was also invited to receive 6 weeks of sleep coaching during their study participation. Coaching participants completed daily sleep diaries to measure sleep or wake time and sleep quality (secondary outcomes). The sleep coaching outcomes are beyond the scope of this manuscript and will be reported elsewhere.

Randomization and Blinding

Randomization occurred at the work site to avoid treatment contamination between employees at the same work site. Before recruitment, all sites in the company (N=1096 sites) were randomized using stratification by total number of employees (ie, 33rd and 67th percentiles; small ≤53 employees, medium=54-73 employees, and large ≥74 employees). Store locations were randomized using allocation sequences generated before the start of the study [34]. Allocation sequences were concealed from the research personnel involved in allocation until the time of group assignment. Participants were informed of their group assignment following completion of the baseline questionnaires.

Intervention and Control Groups

Calm Intervention Group

Participants assigned to the intervention group were instructed to download the consumer-based mobile meditation app, Calm, and were asked to use it autonomously for at least 10 minutes per day during the 8-week intervention period. Meditation in the Calm app uses mindfulness components [35], breathing techniques, and body scans, all of which are consistent with core mindfulness practices, including mindfulness-based stress reduction (nonjudgmental moment-to-moment awareness [36]), and vipassana (objective observation of physical sensation in the body [37]). The frequency, dose, and timing of engagement with the Calm app, as well as its content and use of features, is entirely self-selected by the user. The Calm app is offered internationally in 7 different languages. The Calm app may be accessed by purchasing the app via a subscription-based service or offered by an employer as a benefit; some content is freely available upon download. In addition to using the Calm app, participants had the option to schedule one synchronous 20-minute web-based coaching session with a Calm app coach in the first week to orient them to the Calm app. Participants received weekly SMS text message reminders to use the Calm app on Sundays at noon during the intervention. For the final survey, participants were able enter a raffle to win 1 of 5 boxes of Calm *swag* (pencils, notepads, book, etc), which were provided by the Calm app.

A sample of participants with elevated baseline scores on the Insomnia Severity Index (ISI; see the Measures section; determined as a score of >10 at baseline assessment) were invited to attend 5 weekly sessions over 6 weeks with a certified sleep coach to help them improve their sleep. Sessions were structured around basic sleep hygiene principles (eg, establishing a regular pattern of sleep, engaging in sleep hygiene practices,

sleep restriction (as appropriate for insomnia symptoms), practicing bedtime mindfulness, and improving the sleep environment). The intent was to randomize individuals with elevated ISI scores to either sleep coaching or Calm app only. However, due to low enrollment, we quickly transitioned to offering sleep coaching to all participants with an elevated ISI. Those who opted into sleep coaching were asked to complete a 2-week sleep diary at the beginning and end of the 6-week period.

Waitlist Control Group

Participants randomized to the waitlist control group were instructed via email to continue with their usual routines during the 8-week assessment period. After week 8, they received access to the Calm app for an additional 8 weeks. Waitlisted participants with elevated baseline ISI scores were invited to participate in sleep coaching after completing their waitlist period. To ensure consistency in measurement, they also completed a sleep diary during the waitlist period.

Measures

All participants in both groups were asked to complete electronic self-report assessments of outcomes every 2 weeks from baseline until the completion of the 8-week study period. The constructs measured and the psychometric properties of the assessments used are described.

Demographics

Demographics and individual characteristics (16 items assessing personal characteristics, such as race, ethnicity, work, and medical status) were collected at baseline.

Mental Health

Mental health was measured using the Depression Anxiety Stress Scale (DASS; DASS-21), a 21-item scale assessing symptoms of depression, anxiety, and stress over the past week [38]. The DASS-21 is the short form of the original 42-item measure by Lovibond and Lovibond [39]. It has demonstrated construct validity and maintains the tripartite factor structure of the original DASS-42, effectively distinguishing among the latent constructs of depression, anxiety, and stress via items measuring low positive affect, physiological hyperarousal, and perceived stress [38]. In general population samples, the DASS-21 has shown adequate internal consistency (Cronbach α of .88, .82, and .90 for the depression, anxiety, and stress subscales, respectively) and good convergent and discriminant validity compared with other measures of depression and anxiety [38,39].

Sleep

Insomnia symptoms were assessed among all participants via the ISI, a 7-item self-report questionnaire assessing insomnia symptoms (eg, difficulty falling and staying asleep) during the past 2 weeks and the distress and impairment associated with the symptoms [40]. Items are rated on a 5-point Likert-type scale; total scale scores are obtained by summing the item ratings. The ISI has demonstrated good internal consistency (Cronbach α=.74 in the validation sample), sensitivity to change, and convergence with both objectively measured sleep disturbance and clinician ratings [40]. Daytime sleepiness symptoms were measured using the Epworth Sleepiness Scale, which includes 8 items assessing recent dozing behavior during routine daytime activities (sitting and reading, in conversation, etc) [41]. Items are rated on a 4-point Likert scale from 0 (*would never doze*) to 3 (*high chance of dozing*). The total scores were obtained by summing the item ratings (range 0-24, with higher scores indicating greater sleepiness). The Epworth Sleepiness Scale has shown high internal consistency (Cronbach α=.7 to .9 in varying populations), demonstrates convergent validity with objective measures of sleepiness and sleep disturbance (ie, sleep latency), and differentiates between clinical and nonclinical sleep populations [41,42].

Resilience

Resilience was measured using the Brief Resilience Scale (BRS), which measures an individual's ability to bounce back and recover from stress [21]. Respondents are asked to rate the extent to which 6 statements related to resilience apply to them, on a 5-point Likert-type scale (ie, from *strongly disagree* to *strongly agree*). The BRS has been validated in college students as well as clinical samples (eg, individuals with chronic medical concerns) with good to excellent internal consistency (Cronbach α ranging from .80 to .91 for the overall scale) [43].

Work Productivity and Impairment

The Work Productivity and Activity Impairment (WPAI) Questionnaire–General Health measure (WPAI general health) is a 6-item scale that measures general physical and mental health–related impairments in work and nonwork activities as well as absenteeism and presenteeism [24]. Respondents were asked about current employment, hours missed due to health problems and other reasons, hours worked, and the degree to which health affected productivity during work and in other nonwork activities in the past 7 days. The 4 outcomes generated from the scale are percent work time missed due to health (absenteeism), percent impairment while working due to health (presenteeism), percent overall work impairment due to health (productivity impairment), and percent nonwork activity impairment due to health (nonwork activity impairment).

Medical Care Visits

At weeks 4 and 8, participants were asked to self-report the number of times they had seen a medical provider in the past 4 weeks.

Productivity Cost Savings

The average amount of money saved per employee (US $) due to the improvements in overall work productivity observed in the Calm intervention group versus waitlist control was computed using the human capital approach (HCA) [44,45], which is one of the most widely used methods for estimating the monetary value associated with productivity losses due to a specific cause (eg, mental and physical health problems) [44]. This approach assumes that 1 hour of differential productivity (ie, productivity gained via the Calm app) is equivalent in value to an individual's wages for that same time. In this case, because the WPAI assesses work productivity and impairment over the past 7 days, the HCA produces a weekly estimate of productivity costs associated with employee impairment. Work impairment and associated productivity costs for the present analysis were

derived according to employee wage type—hourly versus salaried. First, the total monetary value of an individual's *potential* productive hours was computed. For hourly workers, the weekly wage was calculated as an employee's hourly wage multiplied by the number of hours per week they worked, and for salaried employees, weekly wages were calculated as an employee's annual salary divided by 52 weeks in a year (given that weekly wages among salaried employees do not depend on the number of hours worked each week). Consistent with HCA, employees' weekly overall work impairment percentages, as indicated by the WPAI overall work impairment metric (ie, percentage of time, relative to the hours worked per week, that an employee is absent or reporting impaired productivity due to mental or physical health problems), were multiplied by their weekly wages to obtain an overall cost of absenteeism and impaired productivity for each individual. These weekly productivity cost values associated with work impairment were then included as the outcome variable in analyses via mixed effects modeling (see the Data Analyses section).

App Use

App use data were provided by the Calm app. Use over time was measured as the average number of sessions and minutes per week of use per employee. Use assessments included measures of overall app use (any component) and the use of specific app components (eg, meditation, music, and sleep stories).

Data Analyses

Power analyses were conducted using G*Power 3.0. Consistent with prior research that has tested similar interventions and measured changes in mental health, resilience, and sleep [46-49], we assumed small to moderate effect sizes for improvements (conservative estimate of Cohen *d*=0.12), thus estimating a total needed sample size of 364. Anticipating approximately 30% attrition, we aimed to enroll a minimum of 500 participants.

All analyses were performed using SPSS (version 27.0; IBM Corp). Baseline comparisons were made between the Calm intervention and waitlist control groups using independent samples 2-tailed *t* tests and Pearson chi-square analyses, as appropriate. As many baseline variables included numerous categories with potentially small cell sizes for chi-square analyses, omnibus baseline group comparisons were conducted with binary indicator variables collapsed across categories (White race vs non-White, presence vs absence of chronic sleep condition, etc). Similar comparisons were made for complete cases (ie, those with data available at all 5 time points from baseline to week 8) versus incomplete cases.

Inferential analyses were conducted using both complete-case (CC; analysis with data from participants completing all assessments) and intent-to-treat (ITT) approaches [50]. ITT analyses were conducted using the mixed models applied to all available data [51,52]. Mixed models were used to analyze group differences in outcomes over time. Mixed models are advantageous because they are well suited for longitudinal data with varying levels of missingness across participants. The models were estimated using maximum likelihood estimation procedures and assuming an autoregressive correlation structure.

To allow for the evaluation of nonlinear change, time was treated as a factor, in which time points were individually dummy coded and compared with baseline. All models included dummy-coded indicators of gender; race; ethnicity; education; employee wage type (ie, hourly vs salaried worker); frontline-worker status (ie, working in retail stores or home services); and the presence of mental, physical, and sleep diagnoses as covariates. In addition, to account for potential effects of sleep coaching (ie, for the Calm intervention group, participants with elevated sleep disturbance who opted for sleep coaching) and completion of sleep diaries (ie, for the waitlist participants invited to sleep coaching after completing the waitlist period), we included 2 dummy-coded indicator variables reflecting enrollment in the sleep coaching program and participation in the program (ie, attended at least one coaching session). Of the 101 participants in the Calm intervention group who indicated interest in enrolling in the CCS program, 55 (54%) completed at least one coaching session. Among those who completed at least one coaching session, the median number of sessions attended was 4 (mean 4.1, SD 1.4). All models allowed for random effects of the person and the work site.

To estimate the effect size of the predictors, we calculated Cohen *d* by dividing the unstandardized regression coefficient by the SD of the outcome variable [53,54]. On the basis of recommendations in the study by Cohen [53], absolute *d* values near 0.30 and below were considered to reflect small effects, at or around 0.50 to reflect medium or moderate effects, and values near 0.80 or above were considered to reflect large effects.

Among participants assigned to the Calm group, descriptive statistics were generated to illustrate app use over time (ie, minutes per week using the Calm app and minutes per week using specific app components).

Results

Baseline Participant Characteristics

Of the 1844 individuals screened for eligibility, 1689 (91.59%) were determined to be eligible and consented to participate. In total, 56.9% (585/1029) of participants were at sites randomized to the Calm intervention group, and 43.14% (444/1029) of participants were at sites randomized to the waitlist control. One participant did not provide sufficient data at any time point (including baseline) to be included in the analyses. Of participants who were randomized, 17.3% (101/444) in the Calm intervention group and 19.6% (87/444) in the control group opted for sleep coaching and sleep diaries. In total, 15% (88/585) of participants had CC data in the Calm intervention group (ie, had data available at all 5 time points), and 23.4% (104/444) of control participants had CC data. Participant flow through the full study is depicted in Figure 1.

Demographic and clinical characteristics of the participants in each group are described in Tables 1 and 2, along with group comparisons. The full sample was relatively evenly split between men (474/1024, 46.28%) and women (518/1024, 50.58%), and 3.12% (32/1024) of the participants identified with another gender. Most participants (803/1028, 78.11%) identified their race as White, and 15.46% (159/1028) identified with Hispanic

or Latinx ethnicity. The groups did not differ at baseline in terms of sex, race, ethnicity, college education (yes or no), presence of a diagnosed sleep condition, or hourly versus salaried wage type (all *P*=.11). The waitlist group had a significantly higher proportion of individuals self-reporting a diagnosed mental health condition (χ^2_1=5.5; *P*=.02), and this variable was included as a covariate in the analyses (as planned) to control for this.

There were no significant baseline differences between groups on measures of depression, stress, insomnia, sleepiness, resilience, absenteeism, presenteeism, or work productivity impairment (Table 3). At baseline, waitlist participants had significantly higher anxiety (*P*=.04) and nonwork activity impairment (*P*=.046), and more frequent medical visits in the 4 weeks before study participation (*P*=.03). However, our modeling approach accounted for these baseline group differences using baseline scores as the reference for each subsequent time point evaluated. See Multimedia Appendix 1 for the group means across all time points.

With regard to comparisons of cases with complete data (ie, participants who provided survey data at all 5 study time points) and incomplete cases (ie, participants who completed surveys at fewer than 5 time points), complete cases were more likely to be people of color (χ^2_1=6.2; *P*=.01), more likely to be non-Hispanic (χ^2_1=4.6; *P*=.03), and less likely to have completed college education (χ^2_1=7.5; *P*=.01). Hourly workers were less likely than salaried workers to complete surveys at all 5 time points (χ^2_1=19.3; *P*<.001); however, among salaried employees, complete cases and incomplete cases did not differ with regard to salary (t_{397}=0.13; *P*=.89), and among hourly workers, complete and incomplete cases did not significantly differ with regard to hourly wage (t_{624}=1.26; *P*=.21).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of study participation and analyzed data. *Note that participants may have been ineligible for more than one reason, all of which are reflected here. During the trial, participants could complete a subsequent assessment even if they missed a prior one.

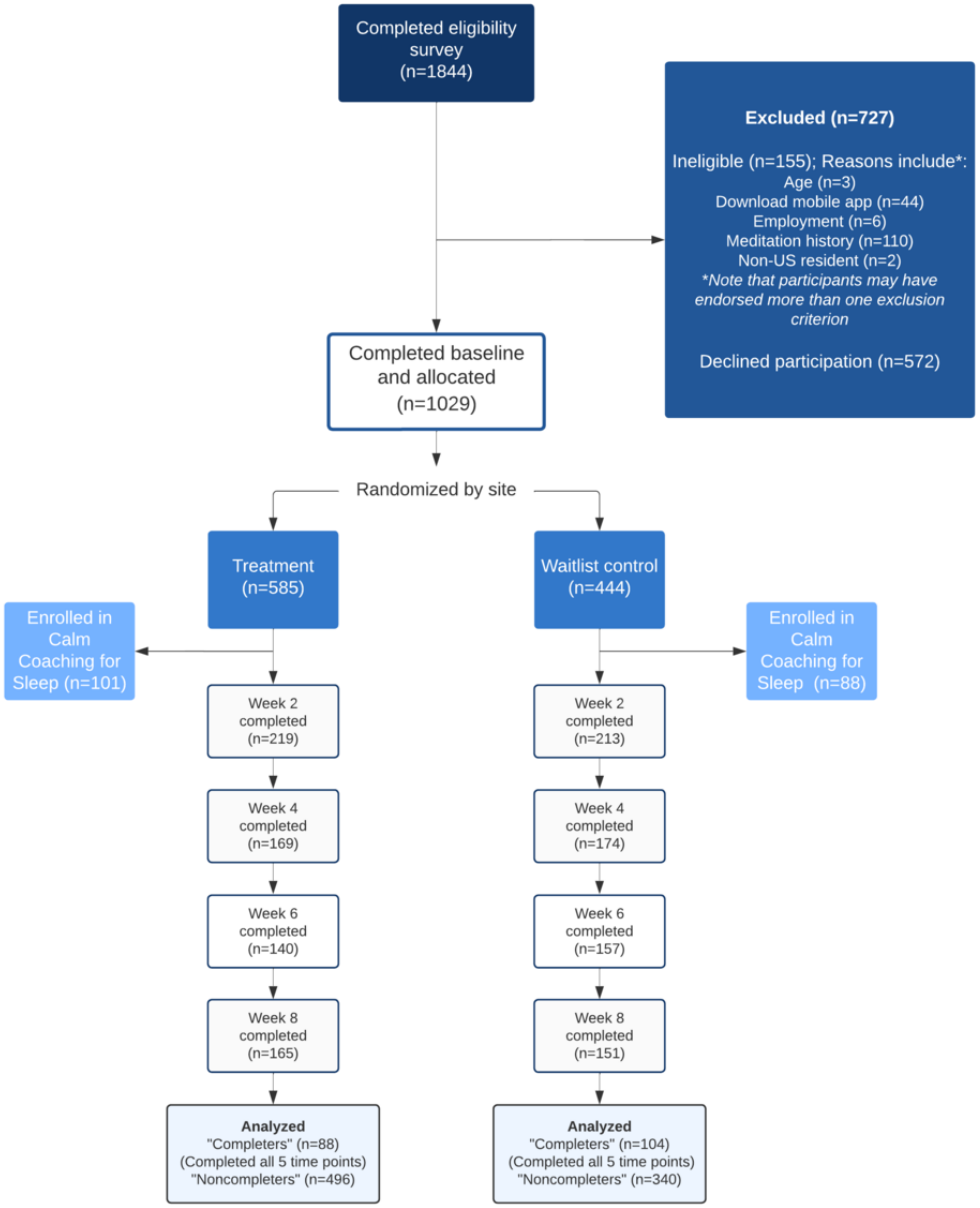


Table 1. Demographic and employment characteristics of the sample[a] (N=1029).

Characteristic	Waitlist, n (%)	Calm, n (%)	Chi-square (*df*)	*P* value
Gender			0.1 (1)	.75
Man	202 (45.7)	272 (46.7)		
Woman	226 (51.1)	292 (50.1)		
Other	14 (3.12)	18 (3.1)		
Race			2.4 (1)	.12
American Indian or Alaskan native	12 (2.7)	16 (2.7)	0.001 (1)	.98
Asian or Asian American	16 (3.6)	33 (5.6)	2.3 (1)	.13
White or European American	340 (77.1)	463 (79.4)	0.8 (1)	.37
Black or African American	26 (5.9)	30 (5.1)	0.3 (1)	.60
Biracial or multiracial	27 (6.1)	26 (4.5)	1.4 (1)	.23
Other	33 (7.45)	20 (3.4)	8.4 (1)	.004
Ethnicity			2.1 (1)	.14
Non-Hispanic or Latino	367 (82.7)	503 (85.9)		
Hispanic or Latino	77 (17.3)	82 (14)		
Education level			0.4 (1)	.53
Less than high school	1 (0.2)	1 (0.1)		
High school diploma	53 (11.9)	52 (89)		
Some college	168 (37.9)	228 (39)		
Associates or 2-year degree	70 (15.8)	60 (10.2)		
Bachelor’s degree	112 (25.2)	175 (29.9)		
Employee type			18.0 (1)	<.001
Salaried (US $ per year)	141 (31.8)	262 (44.8)		
≤39.999	1 (0.7)	0 (0)		
40,000-69,999	30 (21.4)	35 (13.5)		
70,000-99,999	46 (32.9)	89 (34.4)		
100,000-129,999	34 (24.3)	71 (27.4)		
≥130,000	29 (20.7)	64 (24.7)		
Hourly (US $ per hour)	303 (68.2)	323 (55.2)		
<13.00	0 (0)	0 (0)		
13.00-18.99	186 (61.4)	163 (50.5)		
19.00-24.99	87 (28.7)	111 (34.3)		
25.00-30.99	21 (6.9)	35 (10.8)		
≥31.00	9 (2.9)	14 (4.3)		
Employer insurance coverage			0.2 (1)	.60
Yes	316 (71.2)	425 (72.6)		
No	128 (28.8)	160 (27.4)		
Work setting			74.3 (4)	<.001
Market office	3 (0.7)	12 (2.1)		
Store	328 (73.9)	306 (52.3)		
Corporate	87 (19.6)	184 (31.5)		
Home services	2 (0.5)	57 (9.7)		
Other	24 (5.4)	26 (4.4)		

[a]Consistent with the operational definitions of demographic covariates in models of outcomes over time, chi-square tests reflect group comparisons of proportions of men and women, White and racial minority, Hispanic and non-Hispanic, completed and not completed college education, salaried and hourly employment status, and all work setting types.

Table 2. Health characteristics of the sample[a] (N=1029).

Characteristic	Waitlist, n (%)	Calm, n (%)	Chi-square (*df*)	*P* value
Overall physical health			7.7 (4)	.10
Excellent	11 (2.4)	26 (4.4)		
Very good	87 (19.5)	129 (22)		
Good	197 (44.3)	261 (44.6)		
Fair	120 (27)	147 (25.1)		
Poor	29 (6.5)	22 (3.7)		
Chronic health conditions	254 (0.6)	320 (0.5)	1.3 (1)	.26
Mental health condition	175 (41.2)	193 (33.9)	5.5 (1)	.02
Depression	136 (32)	147 (25.8)	4.6 (1)	.03
PTSD[b]	32 (7.5)	35 (6.1)	0.7 (1)	.39
Anxiety disorder	139 (32.7)	139 (24.4)	8.3 (1)	.004
Physical health condition	144 (33.9)	171 (30.1)	1.7 (1)	.20
High blood pressure	71 (16.7)	68 (11.9)	4.6 (1)	.03
High cholesterol	40 (9.4)	41 (7.2)	1.6 (1)	.21
Diabetes	23 (5.4)	18 (3.1)	3.1 (1)	.08
Asthma or other lung disease	54 (12.7)	60 (10.5)	1.1 (1)	.29
Heart disease	5 (1.1)	9 (1.5)	0.3 (1)	.59
Arthritis or rheumatic disease	22 (5.1)	18 (3.1)	2.6 (1)	.11
Cancer	6 (1.4)	7 (1.2)	0.1 (1)	.80
Other chronic condition	48 (11.3)	65 (11.4)	0.004 (1)	.95
Sleep-related conditions	96 (26.5)	121 (24.6)	0.4 (1)	.53
Insomnia	56 (0.1)	63 (0.1)	1.2 (1)	.27
Sleep apnea	33 (0.1)	50 (0.1)	0.3 (1)	.61
Narcolepsy	0 (0)	1 (0)	0.7 (1)	.39
Restless leg syndrome	18 (0.1)	18 (0.1)	0.9 (1)	.35
Somnambulism	1 (0.003)	2 (0)	0.1 (1)	.75
Night terrors	12 (0.03)	8 (0)	2.6 (1)	.12
Other sleep condition	3 (0.01)	10 (0.1)	2.0 (1)	.15

[a]Consistent with operational definitions of health-related covariates in models of outcomes over time, chi-square tests reflect group comparisons of the proportions of the presence or absence of a chronic health condition and the presence or absence of a sleep-related condition.

[b]PTSD: posttraumatic stress disorder.

Table 3. Group differences in outcomes at baseline.

Measure	Waitlist		Calm		*t* test (*df*)	*P* value
	Values, N	Values, mean (SD)	Values, N	Values, mean (SD)		
ISI[a]	444	11.90 (5.87)	584	11.45 (5.65)	1.23 (1026)	.22
ESS[b]	443	7.10 (4.61)	582	7.26 (4.92)	−0.51 (1023)	.61
DASS-21[c]—depression	443	6.47 (5.12)	583	5.96 (4.93)	1.61 (1024)	.11
DASS-21—anxiety	443	4.96 (3.74)	584	4.49 (3.66)	2.02 (1025)	.04
DASS-21—stress	443	8.01 (4.29)	584	7.65 (4.08)	1.39 (1025)	.17
BRS[d]	443	3.29 (0.88)	583	3.31 (0.80)	−0.31 (1024)	.76
WPAI[e]—absenteeism	415	4.95 (14.35)	559	4.48 (14.52)	0.50 (972)	.62
WPAI—presenteeism	406	30.62 (26.14)	547	28.10 (25.69)	1.49 (951)	.14
WPAI—overall work impairment	404	33.08 (28.44)	546	30.30 (27.93)	1.51 (948)	.13
WPAI—activity impairment	438	35.87 (28.70)	574	32.33 (27.16)	2.00 (1010)	.046
Medical visits	444	0.79 (1.17)	584	0.64 (1.08)	2.20 (1026)	.03
Costs due to impaired productivity (US $)	404	334.81 (347.92)	545	354.84 (378.93)	−0.83 (947)	.41

[a]ISI: Insomnia Severity Index.

[b]ESS: Epworth Sleepiness Scale.

[c]DASS-21: Depression Stress Anxiety Scale 21-item.

[d]BRS: Brief Resilience Scale.

[e]WPAI: Work Productivity and Activity Impairment questionnaire. Medical visits were self-reported by employees indicating the number of times they visited a health care provider in the past 4 weeks. Cost due to impaired productivity was calculated based on self-reported pay and WPAI overall work impairment percentages; this metric is reported in US $.

Calm App Effects on Mental Health, Productivity, and Related Outcomes

Mental Health

In the CC analyses, participants at sites randomized to the Calm intervention group had significantly larger reductions in depression at week 8 than did participants at sites randomized to the waitlist control group (Figure 2; Table 4). In ITT analyses (ie, using all available data, inclusive of complete and incomplete cases [51,52]), participants in the Calm intervention group had significantly larger reductions in depression than participants in the waitlist control group at week 6 (Figure S1 in Multimedia Appendix 2). Among complete cases, Calm intervention group participants had significantly larger reductions in anxiety than the control group participants at weeks 4, 6, and 8; in ITT analyses, the effects of week 8 effects were retained. Similarly, complete cases in the Calm intervention group reported significantly larger reductions in stress at weeks 6 and 8 than did CC participants in the waitlist control group, whereas in the ITT analyses, employees using the Calm app had significantly greater reductions in stress at week 6 than those in the waitlist control group. Across all analyses, the effect sizes for changes in mental health indicated small to medium effects.

Figure 2. Estimated marginal means indicating group differences in changes in mental health outcomes over time among study completers. *P<.05; **P<.01; ***P<.001. DASS-21: Depression Anxiety Stress Scale-21 item.

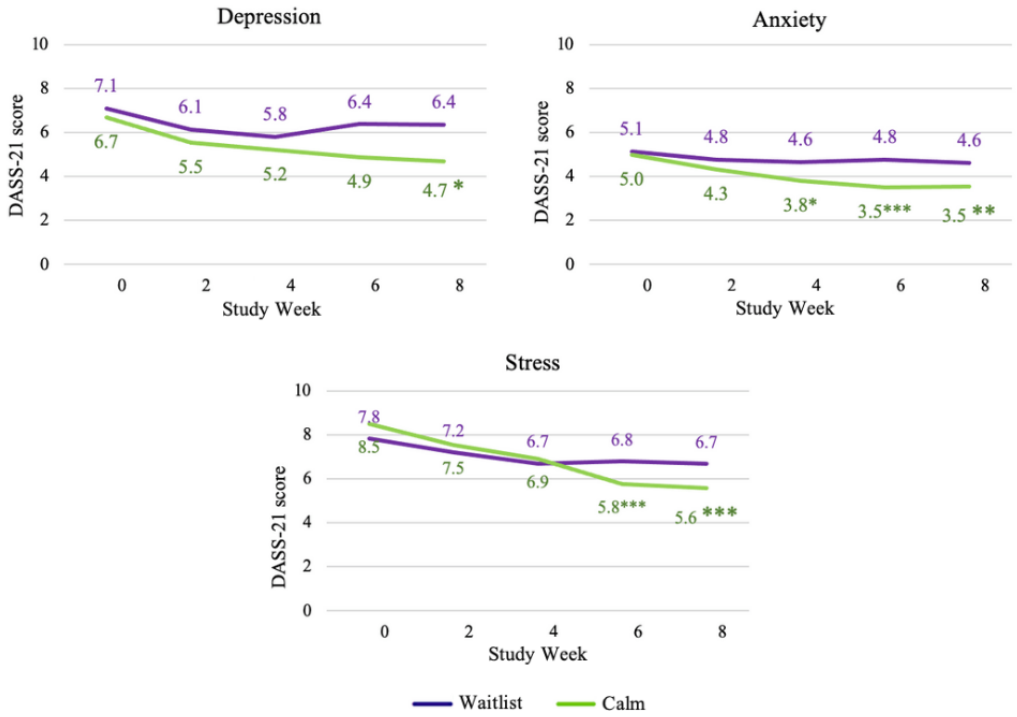


Table 4. Estimates of group differences in changes in mental health over time[a].

Parameter	Complete cases			All available (ITT[b])		
	β (SE; 95% CI)	*P* value	Cohen *d*	β (SE; 95% CI)	*P* value	Cohen *d*
Depression						
Week 2× group	−0.18 (0.57; −1.32 to 0.95)	.75	−0.18	−0.34 (0.30; −0.93 to 0.25)	.26	−0.34
Week 4× group	−0.19 (0.57; −1.31 to 0.93)	.74	−0.10	−0.02 (0.33; −0.67 to 0.62)	.94	−0.02
Week 6× group	−1.09 (0.56; −2.18 to 0.01)	.052	−0.36	−0.73 (0.34; −1.41 to −0.06)	.03	−0.25
Week 8× group	−1.27 (0.55; −2.35 to −0.19)	.02	−0.32	−0.66 (0.34; −1.32 to 0.001)	.05	−0.17
Anxiety						
Week 2× group	−0.28 (0.38; −1.03 to 0.47)	.46	−0.28	−0.02 (0.23; −0.47 to 0.44)	.95	−0.02
Week 4× group	−0.71 (0.34; −1.39 to −0.03)	.04	−0.36	−0.12 (0.24; −0.60 to 0.36)	.63	−0.07
Week 6× group	−1.10 (0.37; −1.83 to −0.37)	.003	−0.37	−0.40 (0.27; −0.94 to 0.13)	.14	−0.14
Week 8× group	−0.92 (0.36; −1.64 to −0.20)	.01	−0.23	−0.57 (0.24; −1.05 to −0.10)	.02	−0.15
Stress						
Week 2× group	−0.36 (0.51; −1.37 to 0.65)	.48	−0.36	−0.14 (0.29; −0.71 to 0.42)	.62	−1.41
Week 4× group	−0.47 (0.47; −1.39 to 0.46)	.32	−0.24	.08 (0.29; −0.49 to 0.65)	.79	−0.08
Week 6× group	−1.70 (0.49; −2.66 to −0.74)	.001	−0.57	−0.77 (0.34; −1.45 to −0.09)	.03	0.02
Week 8×group	−1.78 (0.50; −2.76 to −0.79)	<.001	−0.45	−0.51 (0.33; −1.16 to 0.13)	.12	−0.20

[a]Depression, anxiety, and stress were measured using the Depression Anxiety Stress Scale-21. Complete cases were defined as participants who provided survey data at all 5 time points. All available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For parameter estimates for the complete model, see Tables S1 and S2 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Sleep

In the CC analyses, participants using the Calm app had significantly larger reductions in insomnia symptoms at weeks 4, 6, and 8 than waitlist participants (Figure 3; Table 5). The results from the ITT analysis were similar, with participants using the Calm app reporting greater reductions in insomnia symptoms at weeks 2, 4, 6, and 8 (Figure S2 in Multimedia Appendix 2). Observed effects for insomnia were large in the CC analyses and medium in the ITT analyses. For both CC and ITT analyses, participants using the Calm app also reported greater reductions in daytime sleepiness than waitlist controls at weeks 6 and 8, with small to medium effect sizes observed in CC and small effect sizes observed in ITT analyses.

Figure 3. Estimated marginal means indicating group differences in changes in sleep outcomes over time among study completers. *P<.05; **P<.01; ***P<.001. ESS: Epworth Sleepiness Scale; ISI: Insomnia Severity Index.

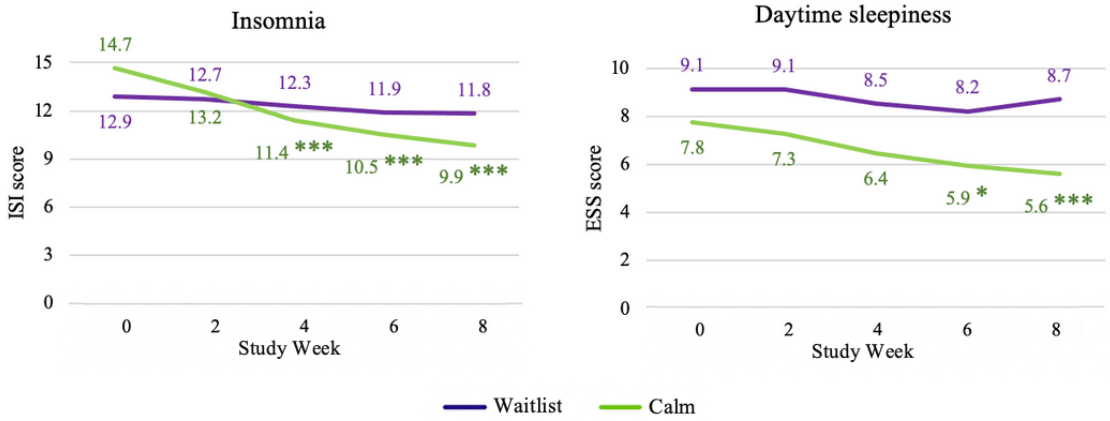


Table 5. Estimates of group differences in changes in sleep symptoms over time[a].

Parameter	Complete cases			All available (ITT[b])		
	β (SE; 95% CI)	*P* value	Cohen *d*	β (SE; 95% CI)	*P* value	Cohen *d*
Insomnia symptoms						
Week 2× group	−1.37 (0.74; −2.84 to 0.09)	.07	−1.37	−0.78 (0.40; −1.57 to −0.005)	.049	−0.02
Week 4× group	−2.65 (0.66; −3.95 to −1.35)	<.001	−1.33	−1.51 (0.39; −2.27 to −0.75)	<.001	−0.37
Week 6× group	−3.18 (0.67; −4.49 to −1.87)	<.001	−1.06	−1.99 (0.43; −2.83 to −1.15)	<.001	−0.22
Week 8× group	−3.74 (0.70; −5.13 to −2.36)	<.001	−0.94	−1.94 (0.48; −2.89 to −1.00)	<.001	1.08
Daytime sleepiness						
Week 2× group	−0.47 (0.51; −1.49 to 0.54)	.36	−0.47	−0.53 (0.31; −1.15 to 0.08)	.09	−0.53
Week 4× group	−0.70 (0.45; −1.60 to 0.19)	.12	−0.35	−0.57 (0.32; −1.21 to 0.06)	.08	−0.29
Week 6× group	−0.91 (0.46; −1.81 to −0.01)	.047	−0.30	−0.77 (0.34; −1.44 to −0.10)	.02	−0.26
Week 8×group	−1.73 (0.47; −2.66 to −0.79)	<.001	−0.43	−1.25 (0.34; −1.92 to −0.58)	<.001	−0.32

[a]Insomnia symptoms were measured using the Insomnia Severity Scale; daytime sleepiness was measured using the Epworth Sleepiness Scale. Complete cases were defined as participants who provided survey data at all 5 time points; all available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For the parameter estimates for the complete model, see Tables S3 and S4 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Resilience

In the CC analysis, Calm intervention group participants had significantly greater (ie, more favorable) resilience scores (BRS) in weeks 4 and 8 (Table 6; small effect sizes), but there were no significant differences observed between groups at any time point in the ITT analysis.

Table 6. Estimates of group differences in changes in resilience over time[a].

Parameter	Complete cases			All available (ITT[b])		
	β (SE; 95% CI)	*P* value	Cohen *d*	β (SE; 95% CI)	*P* value	Cohen *d*
Week 2×group	.07 (0.08; −0.09 to 0.23)	.37	0.07	.01 (0.05; −0.09 to 0.11)	.85	0.01
Week 4×group	.20 (0.08; 0.04 to 0.37)	.02	0.10	.07 (0.06; −0.05 to 0.18)	.24	0.04
Week 6×group	.15 (0.08; −0.01 to 0.31)	.07	0.05	.05 (0.06; −0.07 to 0.16)	.40	0.02
Week 8×group	.21 (0.09; 0.04 to 0.38)	.02	0.05	.08 (0.06; −0.04 to 0.21)	.17	0.02

[a]Resilience was measured using the Brief Resilience Scale. Complete cases were defined as participants who provided survey data at all 5 time points; all available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For parameter estimates for the complete model, see Tables S5 and S6 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Work Productivity and Impairment

For absenteeism, no significant differences were observed between the groups at any time point in either the CC analysis or the ITT analysis (Figure 4; Table 7). For presenteeism, Calm intervention group participants were observed to have significantly lower impairment during work time than control participants at weeks 4, 6, and 8 for the CC analysis (Figure 4) but not for ITT. Small to medium effect sizes were observed for weeks 4, 6, and 8 in the CC analysis. In terms of overall work productivity impairment, participants assigned to the Calm intervention group had significantly lower impairment (ie, were more productive) at weeks 4, 6, and 8 in the CC analysis (Figure 5; medium effect sizes); the effects were not significant in ITT (small effect; Figure S3 in Multimedia Appendix 2). A significant benefit of the Calm app was observed on nonwork activity impairment at weeks 4, 6, and 8 in the CC analysis (Figure 5; medium effect sizes) and week 8 in the ITT analysis (small effect; Figure S4 in Multimedia Appendix 2).

Figure 4. Estimated marginal means indicating group differences in absenteeism and presenteeism over time among study completers. **P*<.05; ***P*<.01; ****P*<.001. WPAI: work productivity and activity impairment.

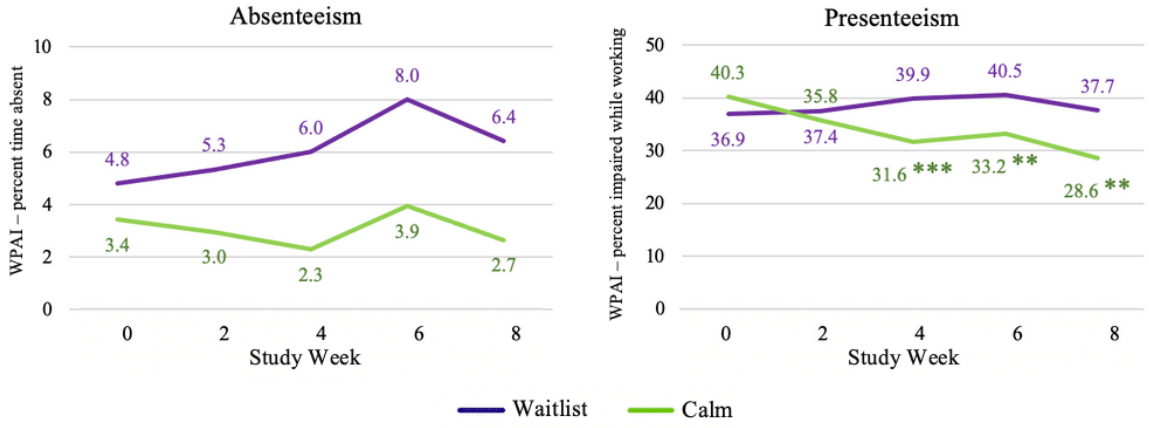


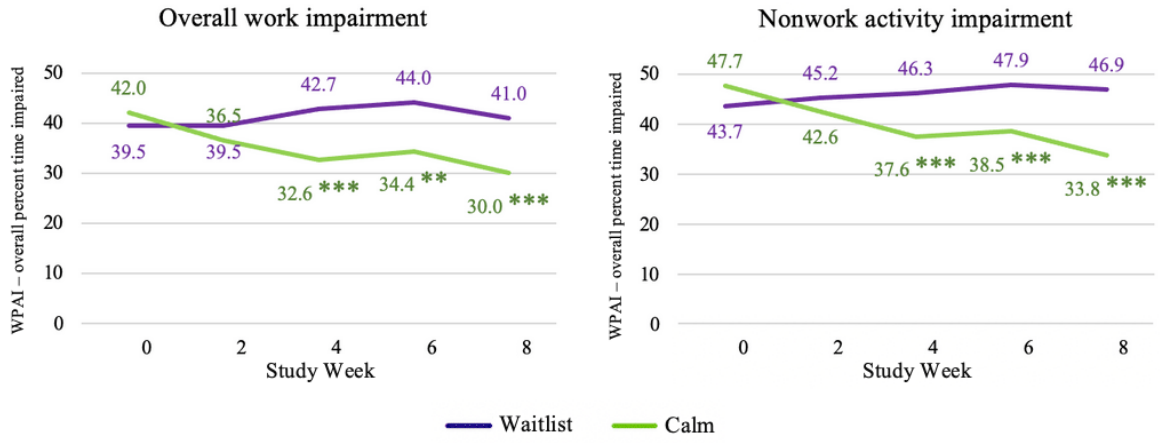
Table 7. Estimates of group differences in changes in work productivity and activity impairment over timea.

Parameter	Complete case			All available (ITT[b])		
	β (SE; 95% CI)	*P* value	Cohen *d*	β (SE; 95% CI)	*P* value	Cohen *d*
Absenteeism						
Week 2× group	−0.99 (2.37; −5.66 to 3.67)	.67	−0.08	−0.19 (1.71; −3.56 to 3.18)	.91	−0.01
Week 4× group	−2.35 (2.04; −6.37 to 1.67)	.25	−0.19	−0.10 (1.48; −3.01 to 2.80)	.95	−0.01
Week 6× group	−2.69 (3.36; −9.34 to 3.95)	.42	−0.21	−1.01 (2.41; −5.76 to 3.74)	.67	−0.07
Week 8× group	−2.40 (2.50; −7.34 to 2.53)	.34	−0.19	−0.16 (1.75; −3.61 to 3.29)	.93	−0.01
Presenteeism						
Week 2× group	−4.90 (4.30; −13.36 to 3.57)	.26	−0.20	3.92 (2.53; −1.05 to 8.88)	.12	0.15
Week 4× group	−11.59 (3.92; −19.32 to −3.86)	.004	−0.46	−2.42 (2.47; −7.28 to 2.43)	.33	−0.10
Week 6× group	−10.58 (4.11; −18.69 to −2.46)	.01	−0.42	−2.39 (2.87; −8.03 to 3.26)	.41	−0.10
Week 8×group	−12.35 (4.37; −20.97 to −3.72)	.01	−0.49	−2.44 (2.80; −7.94 to 3.06)	.38	−0.10
Overall work impairment						
Week 2× group	−5.46 (4.46; −14.25 to 3.34)	.22	−0.21	4.32 (2.72; −1.02 to 9.66)	.11	0.15
Week 4× group	−12.62 (4.24; −20.98 to −4.26)	.003	−0.47	−2.06 (2.72; −7.40 to 3.28)	.45	−0.08
Week 6× group	−12.15 (4.38; −20.79 to −3.51)	.01	−0.45	−2.90 (3.06; −8.91 to 3.10)	.34	−0.11
Week 8×group	−13.47 (4.61; −22.57 to −4.37)	.004	−0.51	−2.28 (2.99; −8.15 to 3.59)	.45	−0.09
Nonwork activity impairment						
Week 2× group	−6.68 (4.25; −15.05 to 1.69)	.12	−0.24	2.78 (2.51; −2.15 to 7.71)	.27	0.10
Week 4× group	−12.70 (3.66; −19.92 to −5.49)	.001	−0.45	−1.75 (2.36; −6.39 to 2.89)	.46	−0.06
Week 6× group	−13.40 (4.01; −21.31 to −5.49)	.001	−0.47	−4.10 (2.89; −9.77 to 1.58)	.16	−0.15
Week 8×group	−17.12 (4.03; −25.06 to −9.18)	<.001	−0.61	−5.64 (2.74; −11.01 to −0.26)	.04	−0.20

[a]All the outcomes were measured using the Work Productivity and Activity Impairment Questionnaire. Complete cases were defined as participants who provided survey data at all 5 time points; all available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For parameter estimates for the complete model, see Tables S7 and S8 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Figure 5. Estimated marginal means indicating group differences in changes in work and nonwork activity impairment over time among study completers. **P*<.05; ***P*<.01; ****P*<.001. WPAI: work productivity and activity impairment.



Medical Care Visits

In both the CC and ITT analyses, participants randomized to the Calm intervention group had a significantly lower frequency of visits with a medical provider at week 8 than individuals in the waitlist control group (Table 8; small effect sizes in both models). No significant differences were observed at week 4.

Table 8. Estimates of group differences in changes in health care visits over timea.

Parameter	Complete cases		
	β (SE; 95% CI)	*P* value	Cohen *d*
Complete cases			
Week 4×group	.02 (0.18; −0.34 to 0.37)	.93	0.01
Week 8×group	−1.05 (0.21; −1.46 to −0.65)	<.001	−0.26
All available (ITT [b])			
Week 4×group	.15 (0.11; −0.05 to 0.36)	.14	0.08
Week 8×group	−0.93 (0.13; −1.19 to −0.68)	<.001	−0.23

[a]Health care visits determined by responses to the question "How many times have you seen a medical provider in the last four weeks?" Complete cases were defined as participants who provided survey data at all 5 time points; all available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For parameter estimates for the complete model, see Tables S9 and S10 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Productivity Cost Savings

There was no significant effect of the Calm app on productivity costs in the ITT analysis. Significant productivity cost savings due to the Calm app were found at weeks 4 and 8 in the CC analysis, with medium effect sizes (Figure 6; Table 9). Among CC participants, the estimated average overall weekly productivity cost associated with work impairment (ie, absenteeism and presenteeism combined) for the Calm intervention group was US $334.13 (SE US $45.61) per employee per week at week 8, compared with US $475.78 (SE US $46.75) at baseline; for waitlist controls, costs associated with work impairment were estimated to be US $433.87.67 (SE US $55.47) at week 8 compared with US $417.55 (SE US $55.95) at baseline. This corresponded to a reduction in weekly costs by US $157.97 (SE US $58.37) per employee attributable to the Calm intervention over 8 weeks.

Figure 6. Estimated marginal means indicating group differences in costs of work impairment over time among study completers.

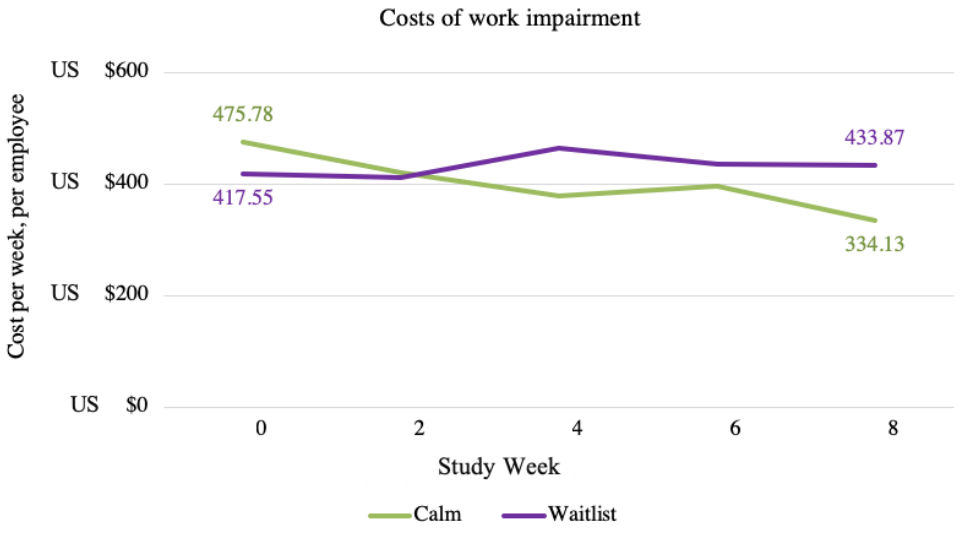


Table 9. Estimates of group differences in changes in work impairment costs over time[a].

Parameter	Complete cases			All available (ITT[b])		
	β (SE; 95% CI)	*P* value	Cohen *d*	β (SE; 95% CI)	*P* value	Cohen *d*
Week 2× group	−49.67 (51.33; −150.98 to 51.65)	.33	−0.15	52.43 (34.85; −16.05 to 120.92)	.13	0.15
Week 4× group	−141.83 (55.87; −251.97 to −31.70)	.01	−0.43	−33.30 (38.21; −108.43 to 41.82)	.38	−0.09
Week 6× group	−97.49 (54.62; −205.20 to 10.21)	.08	−0.29	−24.05 (43.61; −109.87 to 61.76)	.58	−0.07
Week 8×group	−155.66 (57.76; −269.66 to −41.66)	.01	−0.47	−28.41 (40.80; −108.64 to 51.82)	.49	−0.08

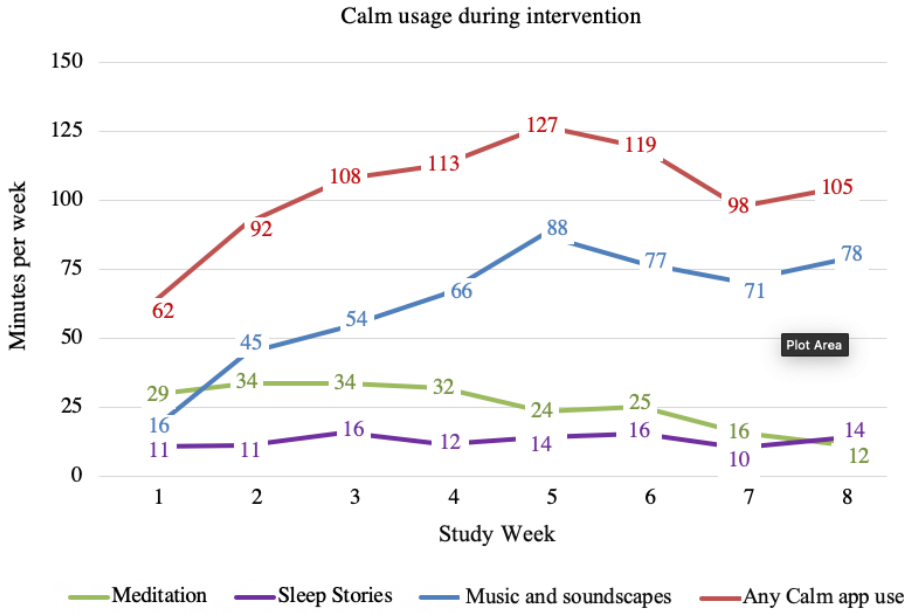
[a]Work impairment costs were calculated by multiplying an employee's weekly pay by their overall work impairment percentage (ie, absenteeism and presenteeism). Complete cases were defined as participants who provided survey data at all 5 time points; all available analyses included all data points from all participants, regardless of the number of survey time points completed. The baseline (week 0) was the reference group for all times by group interaction terms. For parameter estimates for the complete model, see Tables S11 and S12 in Multimedia Appendix 3.

[b]ITT: intent to treat.

Calm App Use

Of the 585 employees randomized to the Calm intervention group, 265 (45.2%) downloaded the Calm app and used it at least once. On average, employees used the Calm app for 102.83 (SD 497.14) minutes per week (average sessions 5.88, SD 23.17). The most popular content was music, soundscapes, and sleep stories (Figure 7; Multimedia Appendix 4).

Figure 7. Average Calm app use per employee per week during the intervention period. This figure depicts the overall app use data (use of any component of the app), as well as the use of the most popular components. Less used components are not included in the figure such that the sum of the components presented does not encompass all app use.



Discussion

Principal Findings

The purpose of this study was to evaluate the experimental effects of the Calm app on employee mental health (depression, anxiety, and stress), sleep, resilience, work productivity outcomes (absenteeism, presenteeism, overall work impairment, and overall activity impairment), and health care use (number of visits with a medical provider) in the workplace. This is the first study to evaluate the Calm app in a workplace setting, specifically with a focus on employee mental health and productivity. Results from a CC analysis (ie, participants completing assessments at all time points) indicated that the Calm app conferred significant benefits to employees in terms of mental health (depression, anxiety, and stress), sleep, resilience, several aspects of productivity (ie, presenteeism, overall work productivity impairment, and nonwork activity impairment), and the frequency of medical visits by the end of the study period. Overall, the effect size calculations indicated small to medium effects. Results from a more conservative analysis with all available data (including incomplete cases that did not provide data at all time points) showed a similar pattern of findings with significant effects observed for depression, anxiety, sleep, nonwork activity impairment, and medical visits.

Mental Health, Sleep, and Resilience

Differences between the Calm group and waitlist control group were consistently observed by the end of the 8-week intervention for depression, anxiety, and stress, indicating that the beneficial effects of mindfulness training on mental health may be cumulative over time. The results align with prior findings,

suggesting that mindfulness interventions provide benefits for anxiety and stress in the workplace and provide further support for their utility in improving depressive symptoms [18,20,22]. Furthermore, these results support the use of digital, app-based interventions as a viable and effective alternative to in-person approaches.

Employees at sites assigned to the Calm intervention group had significant improvements in sleep relative to waitlist control, with robust effects observed in both CC and ITT analyses. Thus, the Calm app provides benefits for mental health and sleep. These effects may have been synergistic, as sleep and mental health are known to have bidirectional effects on one another [55,56]. Indeed, prior work from our team has demonstrated that individuals with sleep disturbance experience improvements in both sleep quality and mental health because of the use of the Calm app [46,57]. In use analyses from this study, participants were observed to use a wide range of Calm app content, including but not limited to sleep stories and other content designed specifically to improve sleep. Although beyond the scope of this study, future work may explore whether engagement with a particular type of content is differentially associated with improvements in certain mental health and sleep constructs.

Relative to the waitlist control, significantly greater improvements in resilience were also observed in the Calm intervention group relative to waitlist control. This effect was only observed in the analysis of CC data and not when incomplete cases were included. Results from the CC analysis align with meta-analytic findings that mindfulness-based interventions in the workplace (inclusive of both in-person and digital delivery modes) improve resilience among employees across a range of employment settings [22].

In the CC analysis, significant reductions were observed in on-the-job work impairment due to physical or mental health concerns (presenteeism) in the Calm intervention group compared with the waitlist control group. However, there were no significant reductions in absenteeism (missed work time) in either the CC or ITT analyses. Thus, it appears that the Calm app may have been more beneficial in terms of improving employees' ability to focus and stay present on the job versus preventing them from missing time from work entirely. These results are consistent with a handful of other studies that have suggested that mindfulness may pose greater benefits for presenteeism and overall productivity than absenteeism [25,26,58]. It is important to note that we also observed a relatively small degree of absenteeism at baseline (approximately 5%); thus, the results may reflect a floor effect with little room for improvement in a general sample. Furthermore, given that absenteeism is typically associated with more severe mental health problems, it is possible that employees require additional support beyond that of a stand-alone app (eg, brief, focused behavioral health coaching) to reduce the time missed when it is caused by mental health problems. Future studies are warranted to test the potential utility of additional support for those with higher levels of absenteeism (especially when absenteeism is attributable to chronic mental health problems).

We also observed significant reductions in the frequency of medical visits among Calm intervention group participants relative to waitlist controls. Given the strong associations among mental health, sleep, and health care use [11-13], it is possible that participants who experienced improvements in mental health because of using the Calm app also felt more able to cope with the stressors they encountered in daily life and work and consequently had a reduced need for medical care, including mental health–related visits. Although this area is less well studied, the findings of this study align with results from previous smaller studies of mindfulness-based interventions in the workplace, which have found that mental and physical health care use and costs are reduced among employees after receiving in-person mindfulness training [59,60].

A key question for employers is whether the benefits of the Calm app on mental health, sleep, and productivity also correspond to financial benefits in terms of cost savings for each employee. To our knowledge, this is the first study to evaluate the effects of a mindfulness app on workplace costs related to health-related work impairment. Although no significant effects were found in the ITT analysis, results from the CC analysis indicated that the Calm app reduced weekly costs by US $155.82 over 8 weeks of the intervention. This is consistent with a recently published study reporting that an employer-sponsored mental health program in the workplace was associated with a positive financial return on investment for employers across multiple worksites and industries [61]. However, this study estimated the cost for mental health services by employees, employers, or insurers and used arbitrary salaries for their analysis. Long-term studies are needed to understand how patterns of productivity cost savings evolve with longer-term employee use of the Calm app and how the results might apply to a broader population with more variable patterns of engagement.

Strengths

This study builds upon the extant literature by improving our understanding of the concurrent effects of the Calm mindfulness app on both mental health outcomes and employee productivity in the workplace using a pragmatic implementation approach that maximizes the generalizability of findings. The Calm app was implemented with a large employer composed of hundreds of sites distributed geographically across the United States and included participants across a variety of income levels and educational backgrounds. The app was deployed entirely by the employer within its workforce, helping to provide pragmatically driven, real-world estimates of the potential reach and uptake of mindfulness apps in similar workplace settings. Participants were blinded to the Calm app brand and received very limited incentive for participating; thus, we anticipate that intervention uptake would be even higher in the future with brand recognition and more employee incentives for using the app as a part of a wellness program [62]. Furthermore, participants were instructed to use the Calm app autonomously and not specifically required to use certain features of the app, which is consistent with how any paying subscribers would engage with the app. Engagement with the app in this study is therefore representative of what employees would likely experience if the Calm app were offered as a component of employee wellness offerings. Finally, this

study was strengthened by the frequent assessment of outcomes *throughout* the intervention period, allowing for a more precise evaluation of the timing at which mindfulness interventions begin to confer their benefits on employee mental health and productivity.

Limitations

Despite this study's strengths, its important limitations must be considered. First, relatively high attrition rates were observed in this study, especially compared with in-person mindfulness interventions in the workplace [27,63,64]. The most common self-reported reason for dropping out (of a total of 63 participants completing the dropout survey) was a lack of time (19/63, 30%) followed by a lack of motivation (11/63, 17%). This high attrition is attributable to two factors: (1) relative to in-person interventions, higher rates of attrition are consistently observed when evaluating digital behavioral health interventions, particularly in the context of pragmatic research where researcher contact does not serve as an incentive or reminder to use and engage with the intervention. Thus, while attrition may have attenuated the positive impact of the Calm app in this sample, the attrition experienced was not unique to this setting or specific intervention. High attrition also precluded us from randomizing participants with sleep disturbance to the sleep coaching intervention. However, in retrospect, one could argue that offering the sleep coaching program to all individuals within the Calm intervention group allowed for a more ecologically valid evaluation, as this is the approach an employer would likely take when the Calm app is offered alongside a more intensive intervention for individuals with elevated levels of distress. It is also important to note that results from an ITT analysis that included all available data from complete cases (those who completed all 5 study assessments) and incomplete cases (those who were lost to attrition or missed assessments) indicated less robust benefits of the Calm app. This is unsurprising given that those participants who started in the Calm intervention group but did not continue using the Calm app would be less likely to benefit. In part due to the sheer size and multisite nature of the workplace in which the Calm app was implemented, we encountered technical and implementation difficulties related to notifying employees of the availability of the study and in sending email reminders to participants to complete assessments. While a majority of the recruitment was conducted through the company's human resources and via their internal email communications, the employer notified us of a potentially low email read rate for messages from their human resources department. This was particularly likely among the large portion of employees for whom regular email checking was not required for their job role (ie, storefront workers); this limited our reach with recruitment. Once individuals were enrolled in the study, we encountered an additional communication barrier, as the company's internet security systems blocked a majority of the study's communication efforts and data collection links or required employees to take extra steps to navigate around security filters. With this in mind, we asked participants to remember to open their emails and complete their assessments outside of their work site and on a personal device not owned by the company, which added an additional step for communication and limited our ability to reach all participants consistently. Implementation of workplace wellness programs is most successful when support for and communication about the intervention comes from within the organization and when messaging is consistent over time [65]. Third, despite music in the Calm app being the most popular content used on the app and the fact that employees used music in the Calm music for an average of 61.72 (SD 23.1) minutes per week, it is uncertain whether listening to music in the Calm app specifically led to improvements in mental health or productivity. Considering the popularity of music in the Calm app and many consumer-based mobile apps that offer similar music content, future studies are warranted to determine the effects of music on mental health and productivity. Furthermore, as is to be expected, we observed variable engagement across work sites, which could indicate other site-level characteristics (eg, workplace culture, networks of communication, and psychological safety within a team) that may have differentially influenced uptake across sites [66,67]. Future pragmatic trials implementing the Calm app in the workplace setting would benefit from additional consideration of these key contextual drivers of reach and uptake. Such trials should focus on testing strategies for engagement among employees with work roles that do not involve regular email communication (SMS text messaging, QR codes in break rooms, social media, etc). Furthermore, future studies should explore organization-level implementation strategies that can maximize the reach and uptake of the intervention.

Conclusions

Commercial apps show promise as a feasible, scalable solution to reduce the burden of mental health problems on employees, improve productivity, and reduce costs for employers. Evidence suggests that mindfulness interventions (including those delivered via a smartphone app) may confer mental health benefits to employees, and the Calm app has been shown to improve mental health in a range of populations. This study adds to this evidence by suggesting that the Calm app improves employees' mental health and productivity. Furthermore, to our knowledge, this is the first study to demonstrate support for productivity cost savings produced by a mindfulness app when implemented pragmatically in a workplace setting.

Acknowledgments

The authors wish to thank Sara Cloonan for assisting with data organization, formatting, and references; Chad Stecher, PhD, for consultation on their analytic approach; and Jehan Sparks, PhD, for consultation and guidance regarding economic and work impairment content.

Conflicts of Interest

JLH and HME-H disclose that they receive an annual salary from the Calm app and hold stock within the company. However, their salary and equity are not dependent on the results of their research. MEP discloses that she is a paid contractor (ie, research coordinator) of the Calm app.

Multimedia Appendix 1
Group means on outcomes over time.
[PDF File (Adobe PDF File), 128 KB - mhealth_v10i9e38903_app1.pdf]

Multimedia Appendix 2
Change over time in intent-to-treat analysis (all available data).
[PDF File (Adobe PDF File), 431 KB - mhealth_v10i9e38903_app2.pdf]

Multimedia Appendix 3
Full treatment effects models.
[PDF File (Adobe PDF File), 177 KB - mhealth_v10i9e38903_app3.pdf]

Multimedia Appendix 4
Employee app use over time.
[PDF File (Adobe PDF File), 35 KB - mhealth_v10i9e38903_app4.pdf]

Multimedia Appendix 5
CONSORT (Consolidated Standards of Reporting Trials) eHealth checklist.
[PDF File (Adobe PDF File), 1163 KB - mhealth_v10i9e38903_app5.pdf]

References

1. Terlizzi EP, Villarroel MA. Symptoms of generalized anxiety disorder among adults: United States, 2019. Centers for Disease Control and Prevention. Hyattsville, MD, USA: National Center for Health Statistics; 2020 Sep. URL: https://www.cdc.gov/nchs/products/databriefs/db378.htm [accessed 2022-04-15]
2. Brody D, Pratt L, Hughes J. Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013–2016. Centers for Disease Control and Prevention. Hyattsville, MD, USA: National Center for Health Statistics; 2018 Feb. URL: https://www.cdc.gov/nchs/products/databriefs/db303.htm/ [accessed 2022-04-15]
3. Cooke JE, Eirich R, Racine N, Madigan S. Prevalence of posttraumatic and general psychological stress during COVID-19: a rapid review and meta-analysis. Psychiatry Res 2020 Oct;292:113347 [FREE Full text] [doi: 10.1016/j.psychres.2020.113347] [Medline: 32763477]
4. Kessler RC. Epidemiology of women and depression. J Affect Disord 2003 Mar;74(1):5-13. [doi: 10.1016/s0165-0327(02)00426-3] [Medline: 12646294]
5. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
6. Ferrie JE, Kumari M, Salo P, Singh-Manoux A, Kivimäki M. Sleep epidemiology--a rapidly growing field. Int J Epidemiol 2011 Dec;40(6):1431-1437 [FREE Full text] [doi: 10.1093/ije/dyr203] [Medline: 22158659]
7. Hammen C, Kim EY, Eberhart NK, Brennan PA. Chronic and acute stress and the prediction of major depression in women. Depress Anxiety 2009;26(8):718-723 [FREE Full text] [doi: 10.1002/da.20571] [Medline: 19496077]
8. Miloyan B, Joseph Bienvenu O, Brilot B, Eaton WW. Adverse life events and the onset of anxiety disorders. Psychiatry Res 2018 Jan;259:488-492 [FREE Full text] [doi: 10.1016/j.psychres.2017.11.027] [Medline: 29154170]
9. Fan LB, Blumenthal JA, Watkins LL, Sherwood A. Work and home stress: associations with anxiety and depression symptoms. Occup Med (Lond) 2015 Mar;65(2):110-116 [FREE Full text] [doi: 10.1093/occmed/kqu181] [Medline: 25589707]
10. Haddadi P, Besharat MA. Resilience, vulnerability and mental health. Procedia Soc Behav Sci 2010;5:639-642. [doi: 10.1016/j.sbspro.2010.07.157]
11. Horenstein A, Heimberg RG. Anxiety disorders and healthcare utilization: a systematic review. Clin Psychol Rev 2020 Nov;81:101894. [doi: 10.1016/j.cpr.2020.101894] [Medline: 32818687]
12. Katon WJ. Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. Biol Psychiatry 2003 Aug 01;54(3):216-226. [doi: 10.1016/s0006-3223(03)00273-7] [Medline: 12893098]

13. Novak M, Mucsi I, Shapiro CM, Rethelyi J, Kopp MS. Increased utilization of health services by insomniacs--an epidemiological perspective. J Psychosom Res 2004 May;56(5):527-536. [doi: 10.1016/j.jpsychores.2004.02.007] [Medline: 15172209]
14. Birnbaum HG, Kessler RC, Kelley D, Ben-Hamadi R, Joish VN, Greenberg PE. Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. Depress Anxiety 2010;27(1):78-89. [doi: 10.1002/da.20580] [Medline: 19569060]
15. Marciniak M, Lage MJ, Landbloom RP, Dunayevich E, Bowman L. Medical and productivity costs of anxiety disorders: case control study. Depress Anxiety 2004;19(2):112-120. [doi: 10.1002/da.10131] [Medline: 15022146]
16. Rosekind MR, Gregory KB, Mallis MM, Brandt SL, Seal B, Lerner D. The cost of poor sleep: workplace productivity loss and associated costs. J Occup Environ Med 2010 Jan;52(1):91-98. [doi: 10.1097/JOM.0b013e3181c78c30] [Medline: 20042880]
17. Evans-Lacko S, Knapp M. Global patterns of workplace productivity for people with depression: absenteeism and presenteeism costs across eight diverse countries. Soc Psychiatry Psychiatr Epidemiol 2016 Nov;51(11):1525-1537 [FREE Full text] [doi: 10.1007/s00127-016-1278-4] [Medline: 27667656]
18. Bartlett L, Martin A, Neil AL, Memish K, Otahal P, Kilpatrick M, et al. A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. J Occup Health Psychol 2019 Mar;24(1):108-126. [doi: 10.1037/ocp0000146] [Medline: 30714811]
19. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Hart R, Eiroa-Orosa FJ. The impact of mindfulness on well-being and performance in the workplace: an inclusive systematic review of the empirical literature. Eur J Work Organ Psychol 2017 Apr 19;26(4):492-513. [doi: 10.1080/1359432x.2017.1308924]
20. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. Mindfulness-based interventions in the workplace: an inclusive systematic review and meta-analysis of their impact upon wellbeing. J Positive Psychol 2019;14(5):625-640. [doi: 10.1080/17439760.2018.1519588]
21. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
22. Vonderlin R, Biermann M, Bohus M, Lyssenko L. Mindfulness-based programs in the workplace: a meta-analysis of randomized controlled trials. Mindfulness 2020 Mar 02;11(7):1579-1598. [doi: 10.1007/s12671-020-01328-3]
23. Sheerin CM, Lind MJ, Brown EA, Gardner CO, Kendler KS, Amstadter AB. The impact of resilience and subsequent stressful life events on MDD and GAD. Depress Anxiety 2018 Feb;35(2):140-147 [FREE Full text] [doi: 10.1002/da.22700] [Medline: 29172241]
24. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993 Nov;4(5):353-365. [doi: 10.2165/00019053-199304050-00006] [Medline: 10146874]
25. Johnson JR, Emmons HC, Rivard RL, Griffin KH, Dusek JA. Resilience training: a pilot study of a mindfulness-based program with depressed healthcare professionals. Explore (NY) 2015;11(6):433-444. [doi: 10.1016/j.explore.2015.08.002] [Medline: 26410675]
26. Bartlett L, Lovell P, Otahal P, Sanderson K. Acceptability, feasibility, and efficacy of a workplace mindfulness program for public sector employees: a pilot randomized controlled trial with informant reports. Mindfulness 2017;8(3):639-654. [doi: 10.1007/s12671-016-0643-4]
27. Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
28. Bartlett L, Martin AJ, Kilpatrick M, Otahal P, Sanderson K, Neil AL. Effects of a mindfulness app on employee stress in an Australian public sector workforce: randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 10;10(2):e30272 [FREE Full text] [doi: 10.2196/30272] [Medline: 35142630]
29. Cherniack M, Lahiri S. Barriers to implementation of workplace health interventions: an economic perspective. J Occup Environ Med 2010 Sep;52(9):934-942. [doi: 10.1097/JOM.0b013e3181f26e59] [Medline: 20798640]
30. Hannon PA, Hammerback K, Garson G, Harris JR, Sopher CJ. Stakeholder perspectives on workplace health promotion: a qualitative study of midsized employers in low-wage industries. Am J Health Promot 2012;27(2):103-110 [FREE Full text] [doi: 10.4278/ajhp.110204-QUAL-51] [Medline: 23113780]
31. Ilvig PM, Bredahl TV, Justesen JB, Jones D, Lundgaard JB, Søgaard K, et al. Attendance barriers experienced by female health care workers voluntarily participating in a multi-component health promotion programme at the workplace. BMC Public Health 2018 Dec 04;18(1):1340 [FREE Full text] [doi: 10.1186/s12889-018-6254-3] [Medline: 30514352]
32. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Mar;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
33. Falcone G, Jerram M. Brain activity in mindfulness depends on experience: a meta-analysis of fMRI studies. Mindfulness 2018 Feb 2;9(5):1319-1329. [doi: 10.1007/s12671-018-0884-5]
34. Haahr M. Random.org: True random number service. RANDOM.ORG. 2022. URL: https://www.random.org [accessed 2021-08-01]

35. Bower JE, Irwin MR. Mind-body therapies and control of inflammatory biology: a descriptive review. Brain Behav Immun 2016 Jan;51:1-11 [FREE Full text] [doi: 10.1016/j.bbi.2015.06.012] [Medline: 26116436]
36. Carmody J, Baer RA. Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. J Behav Med 2008 Mar;31(1):23-33. [doi: 10.1007/s10865-007-9130-7] [Medline: 17899351]
37. Gethin R. The Foundations of Buddhism. Oxford, UK: Oxford University Press; 1998.
38. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
39. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scale. Sydney, Australia: Psychology Foundation; 1995.
40. 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]
41. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [doi: 10.1093/sleep/14.6.540] [Medline: 1798888]
42. Kendzerska TB, Smith PM, Brignardello-Petersen R, Leung RS, Tomlinson GA. Evaluation of the measurement properties of the Epworth sleepiness scale: a systematic review. Sleep Med Rev 2014 Aug;18(4):321-331. [doi: 10.1016/j.smrv.2013.08.002] [Medline: 24135493]
43. Karmiloff-Smith A. Beyond Modularity: A Developmental Perspective on Cognitive Science. Cambridge, MA, USA: MIT Press; 1992.
44. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. Pharmacoeconomics 2004;22(3):165-184. [doi: 10.2165/00019053-200422030-00003] [Medline: 14871164]
45. Ferrari MD. The economic burden of migraine to society. Pharmacoeconomics 1998 Jun;13(6):667-676. [doi: 10.2165/00019053-199813060-00003] [Medline: 10179702]
46. Huberty JL, Green J, Puzia ME, Larkey L, Laird B, Vranceanu AM, et al. Testing a mindfulness meditation mobile app for the treatment of sleep-related symptoms in adults with sleep disturbance: a randomized controlled trial. PLoS One 2021 Jan 7;16(1):e0244717 [FREE Full text] [doi: 10.1371/journal.pone.0244717] [Medline: 33411779]
47. Nakada Y, Sugimoto A, Kadotani H, Yamada N. Verification of effect of sleep health education program in workplace: a quasi-randomized controlled trial. Ind Health 2018 Mar 07;56(1):20-29 [FREE Full text] [doi: 10.2486/indhealth.2017-0022] [Medline: 28855438]
48. Bostock S, Luik A, Espie C. Sleep and productivity benefits of digital cognitive behavioral therapy for insomnia: a randomized controlled trial conducted in the workplace environment. J Occup Environ Med 2016 Jul;58(7):683-689. [doi: 10.1097/JOM.0000000000000778] [Medline: 27257747]
49. Vanhove AJ, Herian MN, Perez AL, Harms PD, Lester PB. Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. J Occup Organ Psychol 2016 Jun;89(2):278-307. [doi: 10.1111/joop.12123]
50. McCoy E. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
51. Chakraborty H, Gu H. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values. Research Triangle Park, NC, USA: RTI Press; 2009.
52. Xi W, Pennell ML, Andridge RR, Paskett ED. Comparison of intent-to-treat analysis strategies for pre-post studies with loss to follow-up. Contemp Clin Trials Commun 2018 Sep;11:20-29 [FREE Full text] [doi: 10.1016/j.conctc.2018.05.008] [Medline: 30023456]
53. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Mahwah, NJ, USA: Lawrence Earlbaum Associates; 1969.
54. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53 [FREE Full text] [doi: 10.1037/a0014699] [Medline: 19271847]
55. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. J Cell Mol Med 2019 Apr;23(4):2324-2332 [FREE Full text] [doi: 10.1111/jcmm.14170] [Medline: 30734486]
56. Yap Y, Slavish DC, Taylor DJ, Bei B, Wiley JF. Bi-directional relations between stress and self-reported and actigraphy-assessed sleep: a daily intensive longitudinal study. Sleep 2020 Mar 12;43(3):zsz250 [FREE Full text] [doi: 10.1093/sleep/zsz250] [Medline: 31608395]
57. Huberty J, Puzia ME, Green J, Vlisides-Henry RD, Larkey L, Irwin MR, et al. A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: analysis from a randomized controlled trial. Gen Hosp Psychiatry 2021;73:30-37 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.004] [Medline: 34537477]
58. Jamieson SD, Tuckey MR. Mindfulness interventions in the workplace: a critique of the current state of the literature. J Occup Health Psychol 2017 Apr;22(2):180-193. [doi: 10.1037/ocp0000048] [Medline: 27643606]
59. Hoge EA, Guidos BM, Mete M, Bui E, Pollack MH, Simon NM, et al. Effects of mindfulness meditation on occupational functioning and health care utilization in individuals with anxiety. J Psychosom Res 2017 Apr;95:7-11. [doi: 10.1016/j.jpsychores.2017.01.011] [Medline: 28314552]

60. Klatt MD, Sieck C, Gascon G, Malarkey W, Huerta T. A healthcare utilization cost comparison between employees receiving a worksite mindfulness or a diet/exercise lifestyle intervention to matched controls 5 years post intervention. Complement Ther Med 2016 Aug;27:139-144. [doi: 10.1016/j.ctim.2016.05.008] [Medline: 27515889]
61. Bondar J, Babich Morrow C, Gueorguieva R, Brown M, Hawrilenko M, Krystal JH, et al. Clinical and financial outcomes associated with a workplace mental health program before and during the COVID-19 pandemic. JAMA Netw Open 2022 Jun 01;5(6):e2216349 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.16349] [Medline: 35679044]
62. Mordecai D, Histon T, Neuwirth E, Heisler WS, Kraft A, Bang Y, et al. How Kaiser permanente created a mental health and wellness digital ecosystem. NEJM Catalyst 2021 Jan 1;2(1):CAT.20.0295. [doi: 10.1056/cat.20.0295]
63. Burton A, Burgess C, Dean S, Koutsopoulou GZ, Hugh-Jones S. How effective are mindfulness-based interventions for reducing stress among healthcare professionals? A systematic review and meta-analysis. Stress Health 2017 Mar;33(1):3-13. [doi: 10.1002/smi.2673] [Medline: 26916333]
64. Montero-Marin J, Kuyken W, Gasión V, Barceló-Soler A, Rojas L, Manrique A, et al. Feasibility and effectiveness of a workplace-adapted mindfulness-based programme to reduce stress in workers at a private sector logistics company: an exploratory mixed methods study. Int J Environ Res Public Health 2020 Mar 03;17(5):1643 [FREE Full text] [doi: 10.3390/ijerph17051643] [Medline: 32138362]
65. Quirk H, Crank H, Carter A, Leahy H, Copeland RJ. Barriers and facilitators to implementing workplace health and wellbeing services in the NHS from the perspective of senior leaders and wellbeing practitioners: a qualitative study. BMC Public Health 2018 Dec 10;18(1):1362 [FREE Full text] [doi: 10.1186/s12889-018-6283-y] [Medline: 30526543]
66. Seward MW, Goldman RE, Linakis SK, Werth P, Roberto CA, Block JP. Showers, culture, and conflict resolution: a qualitative study of employees' perceptions of workplace wellness opportunities. J Occup Environ Med 2019 Oct;61(10):829-835 [FREE Full text] [doi: 10.1097/JOM.0000000000001671] [Medline: 31361680]
67. Puzia ME, Green J, Clarke C, Cloonan S, Huberty J. Examining the associations of using the Calm app with team mindfulness and psychological safety in remote workers. Environ Occup Health Pract 2022;4(1):2022-0001-OA. [doi: 10.1539/eohp.2022-0001-oa]

Abbreviations

BRS: Brief Resilience Scale
CC: complete-case
DASS: Depression Anxiety Stress Scale
HCA: Human Capital Approach
ISI: Insomnia Severity Index
ITT: intent-to-treat
WPAI: Work Productivity and Activity Impairment

Edited by L Buis; submitted 25.04.22; peer-reviewed by J Mascaro, A Teles; comments to author 16.05.22; revised version received 22.07.22; accepted 22.08.22; published 28.09.22.

Please cite as:
Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME
Testing the Pragmatic Effectiveness of a Consumer-Based Mindfulness Mobile App in the Workplace: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(9):e38903
URL: https://mhealth.jmir.org/2022/9/e38903
doi:10.2196/38903
PMID:36169991

Original Paper

The Indirect Effects of a Mindfulness Mobile App on Productivity Through Changes in Sleep Among Retail Employees: Secondary Analysis

Hallie Espel-Huynh[1], PhD; Matthew Baldwin[1], PhD; Megan Puzia[2], MSc; Jennifer Huberty[1,3], PhD

[1]Calm.com, Inc, San Francisco, CA, United States

[2]Behavioral Research and Analytics, LLC, Salt Lake City, UT, United States

[3]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

Corresponding Author:
Hallie Espel-Huynh, PhD
Calm.com, Inc
77 Geary St #3
San Francisco, CA, 94108
United States
Phone: 1 415 984 5864
Email: hallie.espel.huynh@calm.com

Abstract

Background: Chronic sleep disturbance is prevalent among United States employees and associated with costly productivity impairment. Mindfulness interventions improve sleep (ie, insomnia and daytime sleepiness) and productivity outcomes, and mobile apps provide scalable means of intervention delivery. However, few studies have examined the effects of mindfulness mobile apps on employees, and no research to date has tested the role of sleep improvement as a potential mechanism of action for productivity outcomes.

Objective: This study examined the effects of Calm, a consumer-based mindfulness app, and sleep coaching, on productivity impairment among retail employees through the indirect effects of changes in insomnia and daytime sleepiness.

Methods: This study was a secondary analysis of data from a randomized controlled trial (N=1029) comparing the use of Calm (n=585, 56.9%) to a waitlist control (n=444, 43.2%) for 8 weeks among employees of a large retail employer in the United States. A subset of individuals with elevated insomnia symptoms also had access to brief sleep coaching with Calm (n=101, 9.8%). Insomnia symptom severity, daytime sleepiness, and productivity impairment (ie, absenteeism, presenteeism, overall productivity impairment, and non–work activity impairment) were assessed at baseline and weeks 2, 4, 6, and 8. Indirect effects were evaluated with latent growth curve modeling to test whether the Calm intervention (Calm group vs waitlist control) was effective in reducing work productivity impairment through changes in sleep disturbance.

Results: No significant main effects of Calm intervention on productivity impairment were detected for any outcome at α level of .05, with the exception of non–work activity impairment models, in which Calm intervention reduced non–work activity impairment over time (P=.01 and P=.02 for insomnia and sleepiness models, respectively). Significant indirect effects of insomnia were detected for presenteeism (P=.002), overall work productivity (P=.01), and non–work activity impairment (P=.002); Calm intervention produced significantly greater reductions in insomnia symptoms (relative to waitlist control), and decreases in insomnia were associated with decreases in work productivity impairment. There was no significant indirect effect of change in insomnia on changes in absenteeism (P=.20). Furthermore, we detected no significant indirect effects of daytime sleepiness on productivity impairment.

Conclusions: We found that Calm (plus sleep coaching for a small subset of individuals) had beneficial effects on employee sleep, and these benefits on sleep were related to indirect effects on productivity impairment (ie, presenteeism, overall work productivity impairment, and non–work activity impairment). There were no overall main effects of Calm intervention on productivity impairment; however, insomnia appears to be a mechanism associated with benefits for employee productivity. This is one of the first studies to suggest that sleep benefits of a mindfulness mobile app may also indirectly relate to benefits for workplace productivity.

Trial Registration: ClinicalTrials.gov NCT05120310; https://clinicaltrials.gov/ct2/show/NCT05120310

(JMIR Mhealth Uhealth 2022;10(9):e40500) doi:10.2196/40500

KEYWORDS

mindfulness; mobile apps; workforce; workplace; sleep; presenteeism; mobile phone

Introduction

The Problem of Sleep Disturbance

Chronic sleep disturbance is widespread in the United States, with an estimated 36.3% of adults reporting at least one dimension of poor sleep [1]. Insomnia (characterized by difficulty falling asleep, staying asleep, or waking early, along with associated distress or impairment) is a common cause of sleep disturbance linked to substantial public health burden, including increased medical costs, poor mental health, and psychosocial impairment [2,3]. Sleep disturbances such as insomnia also contribute to the daytime effects of poor sleep quality [4], and an estimated one-third of individuals with insomnia experience excessive daytime sleepiness [5]. Insomnia and daytime sleepiness are particularly prevalent among the employed population, with an estimated 19.2% of all workers in the United States experiencing poor sleep quality [6], 23.2% reporting symptoms consistent with insomnia [7], and 16.2% of daytime workers experiencing excessive daytime sleepiness [6]. This paper seeks to summarize a gap in existing research on digital health solutions to address employee sleep and productivity concerns and evaluate the indirect effects of a commercial mindfulness app on employee productivity through sleep improvements.

Both insomnia and daytime sleepiness contribute to costly declines in work productivity (ie, productivity impairment) [6-11]. Compared with their peers with no sleep disturbance, individuals with insomnia experience more absenteeism (missed time from work), presenteeism (time spent being nonproductive at work) [6,7,11], overall work productivity impairment (absenteeism and presenteeism combined), and non–work activity impairment (ie, impairment in functioning outside of work hours) [12]. Daytime sleepiness (because of poor sleep quality or inadequate sleep duration) is associated with increased rates of presenteeism, overall work productivity impairment, and non–work activity impairment [13,14]. Overall, it is estimated that the United States loses an equivalent of 1.23 million working days because of insufficient sleep, which corresponds to approximately 9.9 million working hours [11].

Benefits of Mindfulness in the Workplace

Several studies [15] have demonstrated the beneficial effects of mindfulness-based workplace interventions on employee sleep outcomes, including improved sleep quality, reduced daytime dysfunction, and reduced fatigue [16-18]. Such interventions may serve as an ideal first-line intervention strategy in the workplace, potentially increasing employees' access to help for sleep problems, particularly for those who would not otherwise have access to sleep support [19]. To date, most studies have evaluated the effects of face-to-face interventions, which are costly and complex to implement in workplace settings, may require engagement with highly trained clinicians, and can involve complex payment structures [17,18]. Mobile app interventions have the potential to drastically increase reach and impact [20-23].

There is growing evidence supporting the efficacy of mobile-delivered mindfulness apps on sleep outcomes [20,21]. Various mindfulness apps have been shown to improve sleep in different populations [21], including adults with symptoms of insomnia [24,25], women experiencing environmental stress [21], and parents experiencing pandemic-related stress [26]. Calm is a commercially available mobile mindfulness app that has demonstrated efficacy in improving symptoms of sleep disturbance [24,27]. In addition to Calm's large library of content to support mindfulness practice (eg, guided mindfulness meditations and mindful movements), Calm also includes sleep content that is specifically designed to facilitate improved sleep, including guided sleep meditations, sleep stories for relaxation at bedtime, and sleep-focused music and soundscapes. In a cross-sectional survey of Calm subscribers who used the Calm sleep content, users reported that Calm helped them fall asleep, stay asleep, and get more restful sleep [27]. Moreover, among adults with sleep disturbance, results from a randomized controlled trial indicated that Calm reduced daytime sleepiness, as well as other sleep-related concerns (ie, cognitive and somatic presleep arousal and fatigue), relative to waitlist controls [24].

Gaps in Existing Evidence

To date, evidence for the utility of mobile apps in improving sleep in the workplace is limited. Most workplace evaluations of mindfulness apps have focused solely on mental health outcomes such as anxiety, depression, and psychological well-being [28,29]. Notably, few studies have focused specifically on sleep, especially when examining large-scale pragmatic app implementations across multisite employers [28,29]. Given the widespread prevalence of sleep disturbance in the workforce and its costly impact on productivity, as well as the potential for mindfulness apps to provide a scalable solution for employers, this is a crucial area for further study.

Our team recently evaluated the outcomes of a large randomized controlled trial evaluating Calm's effects on 2 dimensions of sleep disturbance—insomnia and daytime sleepiness (in addition to mental health, resilience, productivity impairment, and health care visits)—among employees at a large retailer in the United States. The results indicated a significant benefit of Calm on insomnia symptoms and daytime sleepiness over the 8-week study period. In a smaller subsample of study completers (but not in the overall sample), significant reductions in employee productivity impairment were observed when examining measures of presenteeism, overall work productivity impairment (absenteeism and presenteeism combined), and non–work activity impairment. No effects were observed for absenteeism in either the full sample or the completer subsample [30].

The extent to which reductions in productivity impairment may occur through reductions in sleep disturbance has yet to be determined. We are aware of the lack of studies examining this

potential link. However, given the strong relationships between sleep and productivity outcomes in the workplace, it is plausible that mindfulness apps may work indirectly on productivity impairment through improvements in sleep. Indeed, our team's earlier work showed that the difference in mental health outcomes observed in individuals who received Calm (plus sleep coaching for a small subset of individuals with elevated insomnia symptoms) relative to the waitlist control occurred through an indirect effect of improvements in sleep [31]. Specifically, the results showed an indirect effect of Calm intervention on improvements in depression and anxiety through cognitive and somatic presleep arousal [31]. Interestingly, in that study, no significant indirect effects were found for daytime sleepiness (insomnia symptoms were not directly measured).

Aims of This Analysis

To date, few studies have examined the effects of a consumer-based mindfulness mobile app on productivity outcomes in the workplace; we are aware of no studies that have assessed whether an app may provide indirect benefits in reducing productivity impairment through the mechanism of reduced sleep disturbance. This study was a secondary analysis of data from the study by Huberty et al [30] to examine whether Calm worked indirectly on productivity impairment (ie, absenteeism, presenteeism, overall productivity impairment, and non–work activity impairment) through the mechanisms of reduced insomnia and daytime sleepiness among employees of a large retailer in the United States.

Methods

Ethics Approval

This study was approved by the institutional review board of Arizona State University (STUDY00014072). All participants completed an electronic informed consent form before participation. The retail employer partner provided input on the study design as it pertained to their interest in supporting employee sleep and well-being as a means of improving productivity and performance.

Participants and Recruitment

The study procedures are described in detail in the study by Huberty et al [30]. Primary outcomes for the clinical trial included insomnia and daytime sleepiness; secondary outcomes included productivity, resilience, depression, anxiety, and stress, as well as sleep diary outcomes (as assessed among individuals offered sleep coaching; see the *Intervention* section for details). Participants were recruited from a large retailer across 1096 work site locations (eg, retail and corporate) distributed throughout the United States between August and December 2021. The number of potentially- eligible employees was approximately 74,000. Recruitment occurred via email invitations distributed by human resources staff and store leaders, as well as on site via flyers posted in workplace break rooms. Email materials and flyers included a QR code and website link that directed participants to a web-based eligibility survey and electronic informed consent form, if eligible. To maintain a pragmatic approach and resemble procedures used if the app were offered as an employee benefit through the retail employer, all recruitment communications were deployed by the employer and not by research personnel. Employees were eligible for the study if they (1) were current employees of the company, (2) were aged at least 18 years, (3) were able to read and understand English, (4) owned a smartphone, (5) were willing to download the Calm app, and (6) had not practiced meditation for >60 minutes per month in the past 6 months [32]. As Calm is a commercially available app with wide name recognizability, potential participants were blinded to Calm during the recruitment process and were informed that the study aimed to test the effects of a *health and wellness app*. After consenting to and completing the baseline self-report measures, participants in the Calm intervention condition received email instructions to download the Calm app at no cost to them.

Participant randomization (Calm intervention vs waitlist control) occurred at the site level (stratified by the number of employees per work site) to reduce possible contamination effects because of the condition among employees at the same work site.

Intervention

Participants in the Calm intervention group were instructed to use Calm autonomously for 10 minutes per day during the 8-week intervention period. To maintain consistency with naturalistic use patterns, no direction was provided regarding the specific content to be used. Weekly reminders to use the app were sent via SMS text messages (to a number specified by the participants); however, no additional communication or incentives were provided for app engagement. The Calm app includes a variety of content types, including guided meditation and breathing exercises, sleep stories, relaxing music, and ambient soundscapes. In addition to the Calm app access, all participants had the option to complete an initial, synchronous, 20-minute *virtual concierge* session with a Calm coach to orient them to the app's offerings and potential areas for behavior change.

A subsample of Calm intervention group participants with elevated insomnia symptoms (≥10 on the Insomnia Severity Index [ISI]; 307/1029, 29.83%; see the *Measures* section) were invited to opt in to receive up to 6 sleep coaching sessions with a trained Calm sleep coach (ie, master's-level behavioral health training and/or coaching certification from an accredited coaching program). Sleep coaching sessions addressed behavioral principles for managing sleep disturbance, including establishing a regular pattern of sleep, engaging in sleep hygiene practices, sleep restriction (as appropriate for insomnia symptoms), practicing bedtime mindfulness, and improving the sleep environment. Participants in the waitlist group were informed that they would receive access to the health and wellness app after completing their final assessment at week 8 (waitlist participants with elevated insomnia symptoms were also invited to participate in sleep coaching alongside access to Calm).

Measures

Overview

Participants were asked to complete self-report assessments via the Qualtrics electronic survey platform every 2 weeks during the study period (baseline and weeks 2, 4, 6, and 8). For each

assessment, participants received an email with a link to complete the web-based survey. Participants who completed the final survey were entered into a raffle for 1 of 5 Calm *swag* prize bags (which included Calm-branded pencils, notepads, and books). No other incentives were provided for the completion of the measures. The assessments included in this secondary analysis are described in this section. Additional outcomes assessed in the full randomized controlled trial included depression, anxiety, perceived stress, resilience, number of medical care visits, and app use (Calm intervention group only).

Baseline Demographic Characteristics

Demographic and individual characteristics (16 items assessing personal characteristics such as race, ethnicity, work, family, and medical history) were collected at baseline.

Insomnia Symptoms

Insomnia symptoms were assessed among all participants using the ISI [33], a 7-item self-report questionnaire assessing insomnia symptoms (eg, difficulty falling and staying asleep) during the past 2 weeks and the distress and impairment associated with the symptoms. Items are rated on a 5-point Likert-type scale, and the total scale scores are obtained by summing the item ratings. A cutoff score of ≥15 indicates moderate to severe insomnia symptoms in the clinical range [33]. The ISI has demonstrated good internal consistency (Cronbach α=.74 in the validation sample), sensitivity to change, and convergence with both objectively measured sleep disturbance and clinician ratings [33].

Daytime Sleepiness

Daytime sleepiness symptoms were measured using the Epworth Sleepiness Scale [34], which includes 8 items assessing recent dozing behavior during routine daytime activities (eg, sitting and reading and conversations). Items are rated on a 4-point Likert scale from 0 (*would never doze*) to 3 (*high chance of dozing*). The total scores are obtained by summing the item ratings (range 0-24, with higher scores indicating greater sleepiness). The Epworth Sleepiness Scale has shown high internal consistency (Cronbach α=.7-.9 in varying populations), demonstrates convergent validity with objective measures of sleepiness and sleep disturbance (ie, sleep latency), and differentiates between clinical and nonclinical sleep populations [34,35].

Productivity Impairment

The Work Productivity and Activity Impairment Questionnaire–General Health measure [36] is a 6-item, well-validated self-report scale that measures general physical and mental health–related impairments in work and nonwork activities, as well as absenteeism and presenteeism. Respondents are asked about current employment, hours missed because of health problems and other reasons, hours worked, and the degree to which health affected productivity during work and in other nonwork activities in the past 7 days. The 4 outcomes generated from the scale are the percentage of work time missed because of health (absenteeism), percentage of impairment while working because of health (presenteeism), percentage of overall work impairment because of health (overall work impairment), and percentage of non–work activity impairment because of health (non–work activity impairment). The Work Productivity and Activity Impairment Questionnaire–General Health has demonstrated sensitivity to productivity impairment in individuals with insomnia [37] and daytime sleepiness [14,37].

Data Analyses

Data Preparation and Sample Size

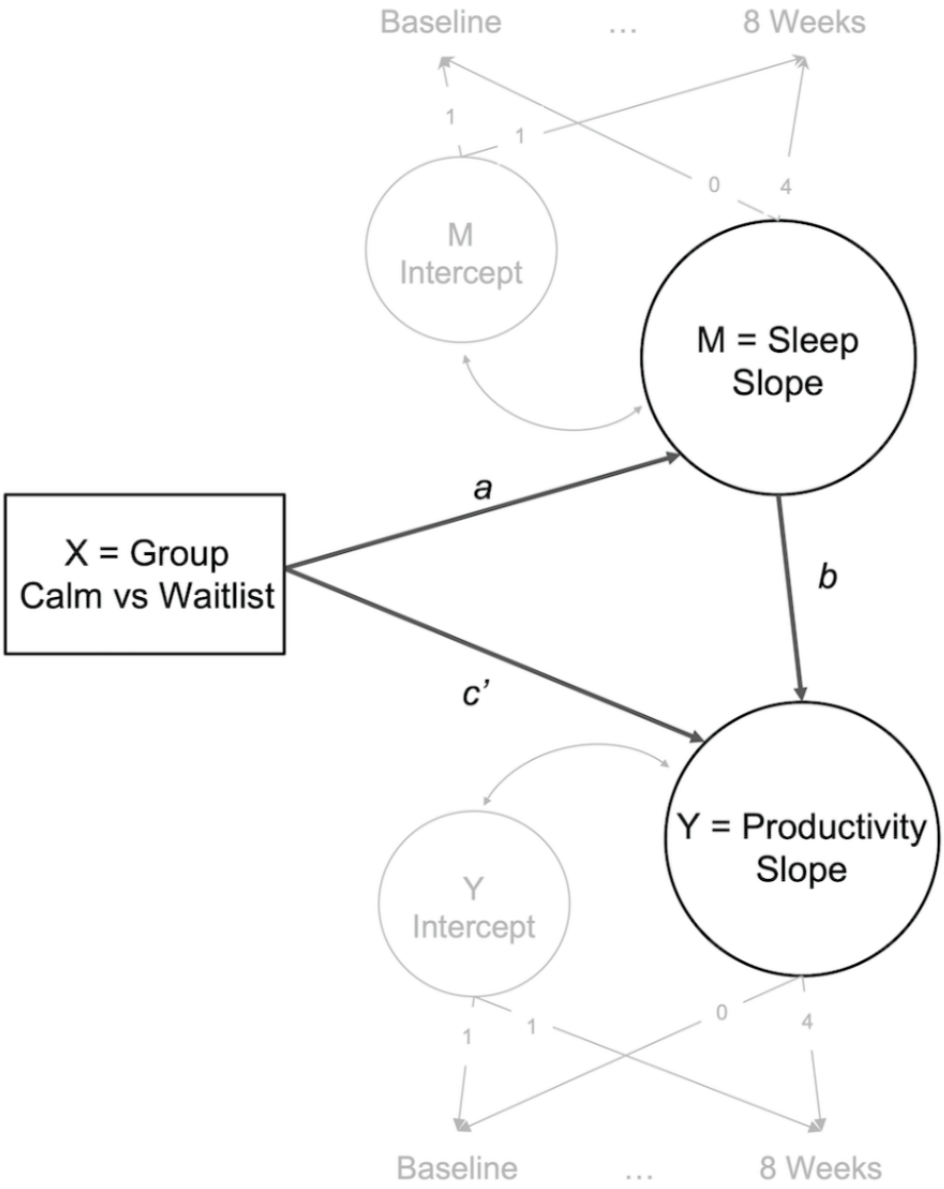
All analyses were conducted using RStudio (build 443; 2022.02.0; version 4.1.2). Participants in the Calm and waitlist control groups were compared in terms of demographic characteristics and baseline sleep- and productivity-related variable scores via 2-tailed independent-sample *t* tests and Pearson chi-square tests as appropriate. In the parent randomized controlled trial, power analyses were conducted using G*Power (version 3.0), yielding a minimum sample size of N=364 to detect a significant effect of Calm on primary outcomes. An initial target sample of N=500 was selected (250 per group), anticipating approximately 30% attrition.

Indirect Effects

Indirect effects were evaluated using latent growth curve models [38] using the *lavaan* package in R [39]. We examined whether the Calm intervention (Calm group vs waitlist control) was effective in improving employee productivity outcomes (ie, latent change in absenteeism, presenteeism, overall productivity impairment, and non–work activity impairment from baseline to week 8) through indirect effects of changes in sleep disturbance (ie, latent change in insomnia and daytime sleepiness from baseline to week 8). Absenteeism percentages were zero-inflated and nonnormally distributed; thus, this variable was recoded as a dichotomous variable, indicating whether an individual reported any missed time during the 1-week before each assessment. Thus, mean level changes (ie, the latent slope) for absenteeism would reflect changes in the likelihood of being absent (vs not absent). All other outcomes were modeled continuously. Covariates were modeled as predictors of both mediator and outcome slopes and intercepts and included gender, hourly worker status, racial minority status, Hispanic ethnicity, presence of a sleep condition, presence of a chronic health condition, and presence of a mental health condition. Separate models were run for each mediator and outcome combination, and latent intercepts and slopes were estimated for both mediators and outcomes. Indicator residual variances were constrained to equality over time, and full information maximum likelihood estimation was used to account for missing data (Multimedia Appendices 1-3 show missing data reports). Regression paths were included for the effect of the intervention on the mediator (*a* path), mediator on the outcome (*b* path), and intervention on the outcome (c' path). The indirect effect was calculated as the product of the *a* and *b* paths [40,41] and was estimated using bias-accelerated and corrected bootstrapping (1000 resamples) [41]. A conceptual model of the analytical approach is shown in Figure 1. Consistent with recommendations for interpreting regression-based indirect effects models with bootstrapping, we did not require a significant total effect before exploration of indirect effects, as evidence suggests that a meaningful indirect effect can be present without the presence of a total effect, and bootstrapping

does not rely on the *causal steps* approach used by traditional stepwise mediation methods [40,42].

Figure 1. Conceptual depiction of the latent growth curve model, with primary elements of the indirect effect in bold. Mean structures, variances, and covariates were removed for simplicity. Factor loadings for baseline and 8 weeks are depicted, although all time points were used to estimate latent intercepts and slopes.



Results

Sample Characteristics

The full trial included 1029 participants (n=585, 56.9% in the Calm group, and n=444, 43.2% in the waitlist control group), 192 (18.7%) of whom completed full assessments at all 5 time points. A description of participant demographics and other baseline characteristics was reported by Huberty et al [30]. Briefly, 50.6% (518/1024) of the participants identified as female, 25.3% (260/1029) identified as belonging to a racial minority group, and 15.5% (159/1029) identified as Hispanic. In terms of clinical characteristics and sleep concerns, 25.97% (267/1028) of the participants scored ≥15 on the ISI, indicating a likely diagnosis of clinical insomnia. Among those who screened positive for insomnia according to their ISI scores, only 32.2% (86/267) indicated that they had been diagnosed with insomnia. Of the included Calm participants, 63.4% (370/584) were screened as eligible for sleep coaching. Of those 370 participants, 101 (27.3%) opted in. In addition, 25.5% (217/851) endorsed the presence of ≥1 diagnosed sleep condition (eg, insomnia, sleep apnea, and restless leg syndrome), and 57.9% (574/992) of the sample indicated that they had been diagnosed with at least one chronic physical health condition (eg, diabetes, hypertension, and chronic asthma). The characteristics of all participants are described in Table 1.

The Calm and waitlist groups did not differ significantly at baseline in terms of gender, race, ethnicity, presence of a chronic medical condition, or presence of a diagnosed sleep condition (including insomnia; all P≥.13; Table 1). The waitlist group had a significantly higher proportion of hourly workers (vs salaried;

χ^2_1=17.4; *P*<.001) and self-reporting of a diagnosed mental health condition (χ^2_1=5.5; *P*=.02). The groups did not differ significantly at baseline in insomnia symptom severity, daytime sleepiness, absenteeism, presenteeism, or overall work productivity impairment (Table 2); however, the waitlist group had significantly greater non–work activity impairment (*P*=.047). Multimedia Appendix 4 shows the group means across all time points. As reported by Huberty et al [30], only 15% (88/585) in the Calm group and 43.2% (192/444) in the waitlist control group provided data for all measures at all time points. In the primary outcome analysis, participants with complete assessment data (ie, all 5 time points) were more likely to identify with a racial minority group (ie, endorsed at least one race other than White), be non-Hispanic, and work on salaried income versus hourly wage [30]. Multimedia Appendices 1-3 show a complete missing data report.

Table 1. Demographic and baseline characteristics of the sample (N=1029).

Characteristic	Waitlist (n=444), n (%)	Calm (n=585), n (%)	Chi-square (*df*)[a]	*P* value
Gender				
Man	202 (45.7)	272 (46.7)	0.1 (1)	.80
Woman	226 (51.1)	292 (50.2)	0.1 (1)	.80
Other	14 (3.2)	18 (3.1)	0.1 (1)	.80
Race				
American Indian or Alaskan Native	12 (2.7)	16 (2.7)	0.0 (1)	.98
Asian or Asian American	16 (3.6)	33 (5.7)	2.3 (1)	.13
White or European American	340 (77.1)	463 (79.4)	0.8 (1)	.37
Black or African American	26 (5.9)	30 (5.2)	0.3 (1)	.60
Biracial or multiracial	27 (6.1)	26 (4.5)	1.4 (1)	.23
Other	33 (7.5)	20 (3.4)	8.4 (1)	.004
Ethnicity				
Not Hispanic or Latino	367 (82.7)	503 (86)	1.9 (1)	.17
Hispanic or Latino	77 (17.3)	82 (14)	1.9 (1)	.17
Employee type				
Salaried	141 (31.8)	262 (44.8)	17.4 (1)[b]	<.001[b]
Hourly	303 (68.2)	323 (55.2)	17.4 (1)[b]	<.001[b]
Insomnia screening (Insomnia Severity Index)				
Moderate or severe insomnia	148 (33.3)	173 (29.6)	1.4 (1)	.23
Unlikely insomnia diagnosis	296 (66.7)	411 (70.4)	1.4 (1)	.23
Chronic health conditions				
At least one	254 (57.2)	320 (56.3)	1.1 (1)	.29
None	170 (38.3)	248 (43.7)	1.1 (1)	.29
Mental health conditions				
At least one	175 (41.3)	193 (34)	5.2 (1)[b]	.02[b]
None	249 (58.7)	375 (66)	5.2 (1)[b]	.02[b]
Sleep-related conditions				
At least one	96 (26.6)	121 (24.7)	0.3 (1)	.58
None	265 (73.4)	369 (75.3)	0.3 (1)	.58

[a]Consistent with operational definitions of demographic covariates in the models of outcomes over time, chi-square tests reflect group comparisons of proportions of male and female, White and racial minority, Hispanic and non-Hispanic, completed and not completed a college education, salaried and hourly employment status, proportions of the likelihood of having insomnia based on the Insomnia Severity Index, presence or absence of a chronic health condition, and the presence or absence of a sleep-related condition.

[b]Significant differences from group comparisons.

Table 2. Baseline scores on outcomes of interest by group (N=1029).

Measure[a]	Waitlist (n=444)		Calm (n=585)		*t* test (*df*)	*P* value
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)		
ISI[b]	444 (100)	11.90 (5.87)	584 (99.8)	11.45 (5.65)	1.22 (1027)	.22
ESS[c]	443 (99.8)	7.10 (4.61)	582 (99.5)	7.26 (4.92)	0.51 (1024)	.61
WPAI[d]: absenteeism	415 (93.5)	0.23 (0.42)	559 (95.6)	0.19 (0.39)	1.77 (973)	.18
WPAI: presenteeism	406 (91.4)	30.62 (26.14)	547 (93.5)	28.10 (25.69)	1.48 (953)	.14
WPAI: overall work impairment	404 (91)	33.08 (28.44)	546 (93.3)	30.30 (27.93)	1.50 (949)	.13
WPAI: activity impairment	438 (98.6)	35.87 (28.70)	574 (98.1)	32.33 (27.16)	1.99[e] (1011)	.047[e]

[a]All measures were assessed continuously, except for absenteeism, which was coded as a dichotomous variable (0=any health-related absence and 1=no absence).

[b]ISI: Insomnia Severity Index.

[c]ESS: Epworth Sleepiness Scale.

[d]WPAI: Work Productivity and Activity Impairment.

[e]Significant differences from group comparisons.

Indirect Effects Analysis Results

The parameter estimates and significance tests for all growth curve models can be found in Multimedia Appendix 5.

Sleep Disturbance and Absenteeism

The model testing the indirect effects of Calm on absenteeism outcomes through insomnia over time showed a good fit. We found a significant beneficial effect of the Calm intervention on changes in insomnia (*a* path) but no significant effect of change in insomnia on change in absenteeism (*b* path). There was no significant direct effect of the Calm intervention on changes in absenteeism, accounting for changes in insomnia (*c′* path), nor was there a significant indirect effect of change in insomnia (*a*×*b* path).

The model testing the indirect effects of Calm on absenteeism outcomes through daytime sleepiness over time also showed a good fit. We found a significant effect of the Calm intervention on changes in daytime sleepiness (*a* path) but no significant effect of change in daytime sleepiness on change in absenteeism (*b* path). There was no significant direct effect of the Calm intervention on absenteeism, accounting for the effects of daytime sleepiness (*c′* path), nor was there a significant indirect effect of change in daytime sleepiness (*a*×*b* path). The model fit and coefficient estimates for both absenteeism models are presented in Table 3.

Table 3. Estimates for models evaluating the indirect effects of Calm on absenteeism through sleep disturbance.

Parameters	Model estimates			
	b (SE)	β	*P* values	95% CI
Insomnia symptoms[a]				
Effect of Calm on Δ insomnia (*a* path; *X*→*M*)	−0.560[b] (0.128[b])	−0.257[b]	<.001[b]	−0.825 to −0.302[b]
Effect of Δ in insomnia on Δ in absenteeism (*b* path; *M*→*Y*)	0.009 (0.006)	.317	.16	−0.003 to 0.021
Total effect of Calm on Δ in absenteeism (*c* path; sum of direct and indirect effects)	0.002 (0.009)	.038	.80	−0.015 to 0.021
Direct effect of Calm on Δ in absenteeism (*c*′ path; *X*→*Y*, accounting for *M*)	0.007 (0.010)	.120	.48	−0.012 to 0.029
Indirect effect of Δ in insomnia (*a*×*b*)	−0.005 (0.004)	−0.081	.20	−0.015 to 0.001
Daytime sleepiness[c]				
Effect of Calm on Δ in daytime sleepiness (*a* path; *X*→*M*)	−0.261[b] (0.092[b])	−0.198[b]	.01[b]	−0.468 to −0.091[b]
Effect of Δ in daytime sleepiness on Δ in absenteeism (*b* path; *M*→*Y*)	−0.011 (0.019)	−0.240	.56	−0.043 to 0.012
Total effect of Calm on Δ in absenteeism (*c* path; sum of direct and indirect effects)	0.002 (0.009)	.032	.82	−0.016 to 0.020
Direct effect of Calm on Δ in absenteeism (*c*′ path; *X*→*Y*, accounting for *M*)	−0.001 (0.010)	.929	.93	−0.015 to 0.017
Indirect effect of Δ in daytime sleepiness (*a*×*b*)	0.003 (0.005)	.047	.60	−0.003 to 0.014

[a]Root mean square error of approximation 0.023, comparative fit index 0.976, and Tucker-Lewis index 0.971.

[b]Significant differences from group comparisons.

[c]Root mean square error of approximation 0.032, comparative fit index 0.952, and Tucker-Lewis index 0.942.

Sleep Disturbance and Presenteeism

The model testing the indirect effects of Calm on presenteeism outcomes through insomnia over time showed an acceptable fit. We found a significant beneficial effect of the Calm intervention on changes in insomnia (*a* path) and a significant effect of change in insomnia on change in presenteeism (*b* path), such that greater reductions in insomnia symptoms were associated with greater reductions in presenteeism (ie, improved productivity). There was no significant direct effect of the Calm intervention on presenteeism, accounting for the effects of insomnia (*c*′ path); however, a significant indirect effect of change in insomnia (*a*×*b* path) was detected. Thus, Calm decreased insomnia, and decreases in insomnia were associated with increases in productivity (decreased presenteeism).

The model testing the indirect effects of Calm on presenteeism through daytime sleepiness over time showed a good fit. We found a significant effect of the Calm intervention on changes in daytime sleepiness (*a* path) but no significant effect of change in daytime sleepiness on change in presenteeism (*b* path). There was no significant direct effect of Calm on presenteeism, accounting for the effects of daytime sleepiness (*c*′ path) and no significant indirect effect of change in daytime sleepiness (*a*×*b* path). Table 4 presents the model fit and coefficient estimates for both presenteeism models.

Table 4. Estimates for models evaluating the indirect effects of Calm on presenteeism through sleep disturbance.

Parameter	Model estimates			
	b (SE)	β	*P* value	95% CI
Insomnia symptoms[a]				
Effect of Calm on Δ in insomnia (*a* path; *X*→*M*)	−0.553[b] (0.125[b])	−0.258[b]	<.001[b]	−0.798 to −0.321[b]
Effect of Δ in insomnia on Δ in presenteeism (*b* path; *M*→*Y*)	3.236[b] (0.798[b])	.821[b]	<.001[b]	2.345 to 4.518[b]
Total effect of Calm on Δ in presenteeism (*c* path; sum of direct and indirect effects)	−0.943 (0.670)	−0.112	.16	−2.368 to 0.329
Direct effect of Calm on Δ in presenteeism (*c*′ path; *X*→*Y*, accounting for *M*)	0.848 (0.740)	.100	.25	−0.547 to 2.215
Indirect effect of Δ in insomnia (*a*×*b*)	−1.791[b] (0.579[b])	−0.212[b]	.002[b]	−2.867 to −0.926[b]
Daytime sleepiness[c]				
Effect of Calm on Δ in daytime sleepiness (*a* path; *X*→*M*)	−0.257[b] (0.094[b])	−0.194[b]	.006[b]	−0.458 to −0.073[b]
Effect of Δ in daytime sleepiness on Δ in presenteeism (*b* path; *M*→*Y*)	1.103 (6.140)	.175	.86	−0.937 to 3.482
Total effect of Calm on Δ in presenteeism (*c* path; sum of direct and indirect effects)	−0.898 (0.675)	−0.108	.18	−2.117 to 0.597
Direct effect of Calm on Δ in presenteeism (*c*′ path; *X*→*Y*, accounting for *M*)	−0.615 (1.837)	−0.074	.74	−1.940 to 1.135
Indirect effect of Δ in daytime sleepiness (*a*×*b*)	−0.284 (1.719)	−0.034	.87	−1.466 to 0.146

[a]Root mean square error of approximation 0.053, comparative fit index 0.911, and Tucker-Lewis index 0.891.

[b]Significant differences from group comparisons.

[c]Root mean square error of approximation 0.038, comparative fit index 0.948, and Tucker-Lewis index 0.937.

Sleep Disturbance and Overall Work Productivity Impairment

The model testing the indirect effect of Calm on overall work impairment outcomes through insomnia over time showed a good fit. We found a significant beneficial effect of the Calm intervention on changes in insomnia (*a* path) and a significant effect of change in insomnia on change in presenteeism (*b* path), such that greater reductions in insomnia were predictive of greater reductions in productivity impairment. There was no significant direct effect of the Calm intervention on overall work impairment, accounting for the effects of insomnia (*c*′ path); however, a significant indirect effect of change in insomnia (*a*×*b* path) was detected. Thus, Calm decreased insomnia, and decreases in insomnia were associated with decreases in productivity impairment.

The model testing the indirect effect of Calm on overall work impairment outcomes through daytime sleepiness over time showed a good fit. We found a significant effect of the Calm intervention on changes in daytime sleepiness (*a* path) but no significant effect of change in daytime sleepiness on change in work impairment (*b* path). There was no significant direct effect of the Calm intervention on work impairment, accounting for the effects of daytime sleepiness (*c*′ path), and no significant indirect (mediating) effect of change in daytime sleepiness (*a*×*b* path). The model fit and coefficient estimates for both models are presented in Table 5.

Table 5. Estimates for models evaluating the indirect effects of Calm on work productivity impairment outcomes through sleep disturbance.

Parameter	Model estimates			
	b (SE)	β	*P* value	95% CI
Insomnia symptoms[a]				
Effect of Calm on Δ in insomnia (*a* path; *X*→*M*)	−0.552[b] (0.127[b])	−0.257[b]	<.001[b]	−0.801 to −0.301[b]
Effect of Δ in insomnia on Δ in overall work impairment (*b* path; *M*→*Y*)	3.129[b] (0.893[b])	.825[b]	<.001[b]	2.185 to 4.534[b]
Total effect of Calm on Δ in overall work impairment (*c* path; sum of direct and indirect effects)	−1.018 (0.693)	−0.125	.14	−2.380 to 0.320
Direct effect of Calm on Δ in overall work impairment (*c*′ path; *X*→*Y*, accounting for *M*)	0.708 (0.855)	.087	.40	−0.700 to 2.203
Indirect effect of Δ in insomnia (*a*×*b*)	−1.726[b] (0.684[b])	−0.212[b]	.01[b]	−2.851 to −0.898[b]
Daytime sleepiness[c]				
Effect of Calm on Δ in daytime sleepiness (*a* path; *X*→*M*)	−0.257[b] (0.095[b])	−0.194[b]	.007[b]	−0.449 to −0.082[b]
Effect of Δ in daytime sleepiness on Δ in overall work impairment (*b* path; *M*→*Y*)	0.978 (9.764)	.162	.92	−0.996 to 3.612
Total effect of Calm on Δ in overall work impairment (*c* path; sum of direct and indirect effects)	−0.971 (0.690)	−0.121	.16	−2.284 to 0.431
Direct effect of Calm on Δ in overall work impairment (*c*′ path; *X*→*Y*, accounting for *M*)	−0.720 (2.295)	−0.090	.75	−2.175 to 1.022
Indirect effect of Δ in daytime sleepiness (*a*×*b*)	−0.251 (2.164)	−0.031	.91	−3.648 to 0.163

[a]Root mean square error of approximation 0.049, comparative fit index 0.921, and Tucker-Lewis index 0.904.

[b]Significant differences from group comparisons.

[c]Root mean square error of approximation 0.039, comparative fit index 0.947, and Tucker-Lewis index 0.935.

Sleep Disturbance and Non–Work Activity Impairment

The model testing the indirect effect of Calm on non–work activity impairment outcomes through insomnia over time showed an acceptable fit. We found a significant beneficial effect of the Calm intervention on changes in insomnia (*a* path) and a significant effect of change in insomnia on change in activity impairment (*b* path), such that greater improvements in insomnia also corresponded to greater reductions in non–work activity impairment (ie, improved functioning). There was no significant direct effect of the Calm intervention on non–work activity impairment, accounting for the effects of insomnia (*c*′ path), although the total effect was significant. A significant indirect effect of change in insomnia (*a*×*b* path) was detected. Thus, Calm decreased insomnia, and decreases in insomnia were associated with decreases in non–work activity impairment.

The model testing the indirect effect of Calm on non–work activity impairment outcomes through daytime sleepiness over time showed a good fit. We found a significant effect of the Calm intervention on changes in daytime sleepiness (*a* path) but no significant effect of change in daytime sleepiness on change in activity impairment (*b* path). The Calm intervention had a significant direct effect on activity impairment (*c* path); however, this effect was not significantly mediated by an indirect effect of the change in daytime sleepiness (*a*×*b* path). The model fit and coefficient estimates for both models are presented in Table 6.

Table 6. Estimates for models evaluating sleep disturbance as a mediator of non–work activity impairment outcomes.

Parameter	Model estimates			
	b (SE)	β	*P* value	95% CI
Insomnia symptoms[a]				
Effect of Calm on Δ in insomnia (*a* path; *X*→*M*)	−0.553[b] (0.129[b])	−0.257[b]	<.001[b]	−0.822 to−0.306[b]
Effect of Δ in insomnia on Δ in non–work activity impairment (*b* path; *M*→*Y*)	3.173[b] (0.705[b])	.736[b]	<.001[b]	2.050 to 4.281[b]
Total effect of Calm on Δ in non–work activity impairment (*c* path; sum of direct and indirect effects)	−1.647[b] (0.673[b])	−0.178[b]	.01[b]	−2.933 to −0.176[b]
Direct effect of Calm on Δ in non–work activity impairment (*c*′ path; *X*→*Y*, accounting for *M*)	0.109 (0.754)	.012	.89	−1.360 to 1.521
Indirect effect of Δ in insomnia (*a*×*b*)	−1.756[b] (0.565[b])	−0.384[b]	.002[b]	−2.995 to −0.871[b]
Daytime sleepiness[c]				
Effect of Calm on Δ in daytime sleepiness (*a* path; *X*→*M*)	−0.257[b] (0.094[b])	−0.194[b]	.006[b]	−0.456 to −0.072[b]
Effect of Δ in daytime sleepiness on Δ in non–work activity impairment (*b* path; *M*→*Y*)	1.345 (6.757)	.195	.84	−0.761 to 3.809
Total effect of Calm on Δ in nonwork activity impairment (*c* path; sum of direct and indirect effects)	−1.609[b] (0.691[b])	−0.176[b]	.02[b]	−2.917 to −0.195[b]
Direct effect of Calm on Δ in non–work activity impairment (*c*′ path; *X*→*Y*, accounting for *M*)	−1.263 (2.114)	−0.138	.55	−2.626 to 0.446
Indirect effect of Δ in daytime sleepiness (*a*×*b*)	−0.346 (2.025)	−0.038	.87	−1.374 to 0.135

[a]Root mean square error of approximation 0.058, comparative fit index 0.906, and Tucker-Lewis index 0.885.

[b]Significant differences from group comparisons.

[c]Root mean square error of approximation 0.039, comparative fit index 0.952, and Tucker-Lewis index 0.941.

Discussion

Principal Findings

The primary aim of this study was to examine whether the beneficial effects of a Calm mobile app intervention for sleep problems (namely, insomnia and daytime sleepiness) would result in indirect benefits for productivity (ie, absenteeism, presenteeism, overall productivity impairment, and non–work activity impairment) among employees of a large retailer in the United States. Web-based, synchronous sleep coaching was also provided for a small subset of those with elevated insomnia symptoms). Even in the absence of significant main effects of the complete Calm intervention on productivity outcomes, we found that changes in insomnia indirectly influenced the effects of Calm intervention on presenteeism, overall productivity impairment, and non–work activity impairment over the 8-week study period. For all 3 outcomes, Calm intervention exerted indirect effects on productivity impairment by producing significant reductions in insomnia symptoms over time, which corresponded to greater reductions in productivity impairment over time. There were no significant indirect effects of Calm intervention on absenteeism through insomnia or on any of the 4 productivity impairment outcomes through daytime sleepiness. Therefore, although the Calm intervention was effective in reducing both insomnia symptoms and daytime sleepiness, it would likely be the reductions in insomnia symptoms that would have primary benefits for productivity. This is one of the first studies to suggest that a mindfulness meditation mobile app implemented in the workplace may reduce productivity impairment through the mechanism of improved sleep. Additional research is needed to replicate and confirm these findings using other employee samples. Although the fit statistics suggested that our models provided an acceptable to good fit for the data, replication studies should also explore additional employee retention strategies (eg, incentives for assessment completion) to increase data availability and enhance confidence in our model estimates. Furthermore, future work could focus more specifically on causal relationships between engagement with the app or coaching, sleep, and work productivity over time.

Changes in Insomnia Symptoms as a Mediator of Productivity Outcomes

It is not surprising that a mindfulness mobile app intervention may improve productivity by reducing insomnia symptoms over time. This finding is consistent with the results from a recent randomized controlled trial, in which sleep-disturbed adults who used Calm experienced reductions in cognitive and somatic presleep arousal, which are 2 aspects of sleep that are closely linked to insomnia [24]. A follow-up analysis from that study showed that improvements in presleep arousal (but not daytime sleepiness) also served as a mechanism of action [31] by which Calm improved other dimensions of well-being (ie, anxiety and

depression). Cognitive and somatic presleep arousal refers to the inability to *shut off* one's thoughts (cognitive) and relax physically (somatic) in preparation for sleep. Both are elevated among individuals with insomnia and are closely related to the core symptoms of insomnia disorder [3,43,44]. Notably, approximately one-fourth of the participants screened positive for a likely insomnia diagnosis with at least moderate clinical symptoms, and only approximately one-third of those individuals self-reported the same diagnosis. This suggests that many employees may not be aware of and/or have access to care for their clinical insomnia symptoms. Without tools such as mindfulness apps to address their symptoms, this illustrates a missed opportunity to influence employee sleep and productivity.

We also did not find a significant indirect effect of the Calm intervention on absenteeism through changes in insomnia symptoms. In the available literature, absenteeism has been less consistently related to sleep disturbance in the workplace than to other measures of productivity impairment [7]. We also observed relatively low rates of health-related absenteeism at baseline (ie, only 204/974, 20.9%, of the sample reported an absence at baseline). Thus, the effects of the intervention on absenteeism may be relevant for only a small minority of retail employees.

Employers incur immense costs because of insomnia-related productivity impairments. Our findings suggest that reductions in insomnia symptoms may be a key driver of reduced productivity impairment for employees and that mindfulness apps such as Calm, plus additional coaching support where indicated, can serve as accessible and scalable first-line interventions to address both. Future work could examine how specific content——or use at particular times of day—may affect sleep and productivity. For example, one could examine whether it is more beneficial to use sleep content at bedtime, general mindfulness meditation while winding down in the evening, or a combination of both, as well as whether this varies by employee or work schedule (eg, varying shift work vs consistent schedule). Furthermore, one may examine whether coaching works synergistically with the app itself, such that app use increases after certain coaching interactions.

Change in Daytime Sleepiness as a Mediator of Productivity Outcomes

It is somewhat surprising that the Calm intervention did not appear to work indirectly on productivity impairment through changes in daytime sleepiness, especially given that daytime sleepiness is closely linked to workplace productivity and manifests primarily during work hours. However, this finding is consistent with results from a prior randomized controlled trial evaluating the effects of Calm in adults with sleep disturbance. In that study, Calm reduced symptoms of daytime sleepiness, and these changes were directly associated with (but did not mediate) reductions in anxiety and depression [24]. Daytime sleepiness is also driven by many other sleep concerns, in addition to insomnia and insufficient sleep duration, including medical problems such as obstructive sleep apnea, which are not likely to be improved by meditation [45]. Thus, although commercial mindfulness apps may improve daytime sleepiness in certain cases, this may be a secondary benefit rather than a mechanism of action.

Therefore, the primary benefit of a mindfulness mobile app intervention on productivity impairment appears to be in helping users fall asleep more easily and reduce their distress about sleep (which is a key feature of insomnia), which presumably improves their ability to focus and be productive at work. Mechanistically, mindfulness practice encourages nonattachment to one's internal experience (eg, racing thoughts, tense body, and difficulty settling for bedtime) [46], which may be particularly beneficial when attempting to fall asleep or go back to sleep after nighttime awakening. Taken together, these findings suggest that addressing insomnia symptoms is key to reducing work productivity impairment and non–work activity impairment among retail employees.

Strengths

This study builds on existing knowledge about the benefits of implementing a mindfulness meditation mobile app in a workplace setting and its ability to affect productivity impairment through improvements in sleep. The strengths of this study include (1) a pragmatic randomized trial in a diverse sample of employees of a large retail company with sites distributed across the United States, thus maximizing generalizability to other US workplace settings; (2) a mechanistic approach to enhance our knowledge about *how* mobile mindfulness apps such as Calm, along with supplemental coaching, may reduce productivity impairment in employees; (3) frequent assessment of outcomes (ie, every 2 weeks throughout the study period), which allowed for a more rigorous evaluation of indirect effects; and (4) naturalistic use among participants (ie, participants were instructed to use the app for 10 minutes per day but received no other instructions or curation of content). Thus, employee engagement with Calm (and the observed benefit) was likely representative of how employees would use the app and associated services if they were offered as an employee benefit.

Limitations

Despite its strengths, this study had several limitations. First, there was a large degree of attrition, leading to substantial missingness and potential bias in our parameter estimates (see Multimedia Appendices 1-3 for missing data reports), which may also correspond to a more limited engagement with the Calm intervention. There were multiple barriers to communicating with employees throughout the recruitment and study periods, which contributed to attrition (reported in the study by Huberty et al [30]). We attempted to account for any systematic differences between included and nonincluded cases by including covariates for any demographic characteristics that could be related to differential study attrition and patterns of missingness among outcome variables and covariates. Although data were not missing completely at random, we considered that data could be assumed to be missing at random and that our selected covariates could explain some of the missingness of key outcomes. The full information likelihood estimator, alongside the inclusion of strong auxiliary variables, has been shown to reproduce unbiased estimates of effects, even when missingness is initially nonignorable [47,48]. Considering that

participants in this study shared a large degree of initial data (eg, demographics and baseline measures of sleep and productivity outcomes) and that included covariates were related to missingness, we feel confident that the full information and auxiliary variable approach we used here adequately accounts for missing data. Future studies with more comprehensive implementation strategies are required to overcome some of the barriers encountered with employee communication, participation, and attrition. Second, although participants were diverse in terms of gender, wage type, and the types of experienced medical and mental health concerns, the study sample was limited in terms of racial and ethnic diversity. Thus, future studies with more focused recruitment efforts on employees from diverse racial and ethnic backgrounds are warranted. Third, this study involved the Calm app plus sleep coaching, and it is unclear which effects can be attributed to the Calm app alone versus sleep coaching intervention. Finally, this study is limited by its use of self-report measures, with no formal clinician assessment of sleep disturbances. Future work may incorporate more formal sleep assessment data, for example, by using objective monitoring (ie, actigraphy), direct clinician observation, or the incorporation of employee data from medical claims or electronic health records.

Conclusions

Commercial mindfulness apps provide a unique opportunity to deploy easily accessible and scalable interventions that can improve employees' sleep and productivity. Mindfulness interventions are increasingly being shown to improve sleep among the employed population, and there is growing evidence that mobile-delivered mindfulness apps such as Calm can serve as a feasible alternative to in-person interventions, with similar benefits. The present results show that not only does a mobile mindfulness app produce benefits for sleep among the working population but that improvements in insomnia in particular may play a key mechanistic role in reducing productivity impairment.

Acknowledgments

The authors would like to acknowledge the retail partner and its employees for their collaboration in this project, Breanna Banks for her assistance with references and formatting, Sara Cloonan for assistance with references, and Taylor Neher for her involvement in data collection for the randomized controlled trial.

Conflicts of Interest

HEH, MB, and JH disclose that they receive an annual salary from Calm and hold stock within the company. However, their salary and equity were not dependent on the results of their research. MP discloses that she is a paid contractor (ie, the research coordinator) of Calm.

Multimedia Appendix 1
Missing data patterns by week.
[PDF File (Adobe PDF File), 57 KB - mhealth_v10i9e40500_app1.pdf]

Multimedia Appendix 2
Missing data patterns among sleep variables and covariates.
[PDF File (Adobe PDF File), 121 KB - mhealth_v10i9e40500_app2.pdf]

Multimedia Appendix 3
Missing data patterns among productivity variables and covariates.
[PDF File (Adobe PDF File), 167 KB - mhealth_v10i9e40500_app3.pdf]

Multimedia Appendix 4
Mean scores of outcome measures over time by group (all available data).
[PDF File (Adobe PDF File), 28 KB - mhealth_v10i9e40500_app4.pdf]

Multimedia Appendix 5
Growth curve mediation model: insomnia and absenteeism.
[PDF File (Adobe PDF File), 259 KB - mhealth_v10i9e40500_app5.pdf]

References

1. Roth T, Jaeger S, Jin R, Kalsekar A, Stang PE, Kessler RC. Sleep problems, comorbid mental disorders, and role functioning in the national comorbidity survey replication. Biol Psychiatry 2006 Dec 15;60(12):1364-1371 [FREE Full text] [doi: 10.1016/j.biopsych.2006.05.039] [Medline: 16952333]
2. Mai E, Buysse DJ. Insomnia: prevalence, impact, pathogenesis, differential diagnosis, and evaluation. Sleep Med Clin 2008;3(2):167-174 [FREE Full text] [doi: 10.1016/j.jsmc.2008.02.001] [Medline: 19122760]

3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
4. Shekleton JA, Rogers NL, Rajaratnam SM. Searching for the daytime impairments of primary insomnia. Sleep Med Rev 2010 Feb;14(1):47-60. [doi: 10.1016/j.smrv.2009.06.001] [Medline: 19963414]
5. Ford ES, Cunningham TJ, Giles WH, Croft JB. Trends in insomnia and excessive daytime sleepiness among U.S. adults from 2002 to 2012. Sleep Med 2015 Mar;16(3):372-378 [FREE Full text] [doi: 10.1016/j.sleep.2014.12.008] [Medline: 25747141]
6. Yong LC, Li J, Calvert GM. Sleep-related problems in the US working population: prevalence and association with shiftwork status. Occup Environ Med 2017 Feb;74(2):93-104. [doi: 10.1136/oemed-2016-103638] [Medline: 27609325]
7. Kessler RC, Berglund PA, Coulouvrat C, Hajak G, Roth T, Shahly V, et al. Insomnia and the performance of US workers: results from the America insomnia survey. Sleep 2011 Sep 01;34(9):1161-1171 [FREE Full text] [doi: 10.5665/SLEEP.1230] [Medline: 21886353]
8. Redeker NS, Caruso CC, Hashmi SD, Mullington JM, Grandner M, Morgenthaler TI. Workplace interventions to promote sleep health and an alert, healthy workforce. J Clin Sleep Med 2019 Apr 15;15(4):649-657 [FREE Full text] [doi: 10.5664/jcsm.7734] [Medline: 30952228]
9. Pilcher JJ, Morris DM. Sleep and organizational behavior: implications for workplace productivity and safety. Front Psychol 2020 Jan 31;11:45 [FREE Full text] [doi: 10.3389/fpsyg.2020.00045] [Medline: 32082218]
10. Espie CA, Pawlecki B, Waterfield D, Fitton K, Radocchia M, Luik AI. Insomnia symptoms and their association with workplace productivity: cross-sectional and pre-post intervention analyses from a large multinational manufacturing company. Sleep Health 2018 Jun;4(3):307-312. [doi: 10.1016/j.sleh.2018.03.003] [Medline: 29776626]
11. Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Q 2017 Jan;6(4):11 [FREE Full text] [Medline: 28983434]
12. Roth T, Ancoli-Israel S. Daytime consequences and correlates of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. II. Sleep 1999 May 01;22 Suppl 2:S354-S358. [Medline: 10394607]
13. Ohayon MM, Dauvilliers Y, Reynolds 3rd CF. Operational definitions and algorithms for excessive sleepiness in the general population: implications for DSM-5 nosology. Arch Gen Psychiatry 2012 Jan;69(1):71-79 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.1240] [Medline: 22213791]
14. Dean B, Aguilar D, Shapiro C, Orr WC, Isserman JA, Calimlim B, et al. Impaired health status, daily functioning, and work productivity in adults with excessive sleepiness. J Occup Environ Med 2010 Feb;52(2):144-149. [doi: 10.1097/JOM.0b013e3181c99505] [Medline: 20134351]
15. Bartlett L, Martin A, Neil AL, Memish K, Otahal P, Kilpatrick M, et al. A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. J Occup Health Psychol 2019 Feb;24(1):108-126. [doi: 10.1037/ocp0000146] [Medline: 30714811]
16. Querstret D, Cropley M, Fife-Schaw C. Internet-based instructor-led mindfulness for work-related rumination, fatigue, and sleep: assessing facets of mindfulness as mechanisms of change. A randomized waitlist control trial. J Occup Health Psychol 2017 Apr;22(2):153-169. [doi: 10.1037/ocp0000028] [Medline: 27054503]
17. Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
18. Klatt M, Norre C, Reader B, Yodice L, White S. Mindfulness in motion: a mindfulness-based intervention to reduce stress and enhance quality of sleep in Scandinavian employees. Mindfulness (N Y) 2017;8(2):481-488. [doi: 10.1007/s12671-016-0621-x]
19. Wickwire EM. Value-based sleep in the workplace. Sleep 2016 Oct 01;39(10):1767-1768 [FREE Full text] [doi: 10.5665/sleep.6150] [Medline: 27634808]
20. Jiang A, Rosario M, Stahl S, Gill JM, Rusch HL. The effect of virtual mindfulness-based interventions on sleep quality: a systematic review of randomized controlled trials. Curr Psychiatry Rep 2021 Jul 23;23(9):62 [FREE Full text] [doi: 10.1007/s11920-021-01272-6] [Medline: 34297230]
21. Rung AL, Oral E, Berghammer L, Peters ES. Feasibility and acceptability of a mobile mindfulness meditation intervention among women: intervention study. JMIR Mhealth Uhealth 2020 Jun 02;8(6):e15943 [FREE Full text] [doi: 10.2196/15943] [Medline: 32442147]
22. Jimenez P, Bregenzer A. Integration of eHealth tools in the process of workplace health promotion: proposal for design and implementation. J Med Internet Res 2018 Feb 23;20(2):e65 [FREE Full text] [doi: 10.2196/jmir.8769] [Medline: 29475828]
23. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018 May 10;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
24. Huberty JL, Green J, Puzia ME, Larkey L, Laird B, Vranceanu AM, et al. Testing a mindfulness meditation mobile app for the treatment of sleep-related symptoms in adults with sleep disturbance: a randomized controlled trial. PLoS One 2021 Jan 7;16(1):e0244717 [FREE Full text] [doi: 10.1371/journal.pone.0244717] [Medline: 33411779]

25. Low T, Conduit R, Varma P, Meaklim H, Jackson ML. Treating subclinical and clinical symptoms of insomnia with a mindfulness-based smartphone application: a pilot study. Internet Interv 2020 Sep;21:100335 [FREE Full text] [doi: 10.1016/j.invent.2020.100335] [Medline: 32939342]
26. Burgess A, Cavanagh K, Strauss C, Oliver BR. Headspace for parents: qualitative report investigating the use of a mindfulness-based app for managing parents' stress during COVID-19. BJPsych Open 2021 Dec 17;8(1):e15 [FREE Full text] [doi: 10.1192/bjo.2021.1070] [Medline: 34956647]
27. Huberty J, Puzia ME, Larkey L, Vranceanu A, Irwin MR. Can a meditation app help my sleep? A cross-sectional survey of Calm users. PLoS One 2021 Oct 22;16(10):e0257518 [FREE Full text] [doi: 10.1371/journal.pone.0257518] [Medline: 34679078]
28. Bartlett L, Martin AJ, Kilpatrick M, Otahal P, Sanderson K, Neil AL. Effects of a mindfulness app on employee stress in an Australian public sector workforce: randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 10;10(2):e30272 [FREE Full text] [doi: 10.2196/30272] [Medline: 35142630]
29. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
30. Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME. Bringing calm to work: a randomized controlled trial testing the pragmatic effectiveness of a consumer-based mindfulness mobile app in the workplace. JMIR Mhealth Uhealth (forthcoming) 2022 Aug 22 (forthcoming) [FREE Full text] [doi: 10.2196/38903]
31. Huberty J, Puzia ME, Green J, Vlisides-Henry RD, Larkey L, Irwin MR, et al. A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: analysis from a randomized controlled trial. Gen Hosp Psychiatry 2021;73:30-37 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.004] [Medline: 34537477]
32. Carmody J, Baer RA. Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. J Behav Med 2008 Feb;31(1):23-33. [doi: 10.1007/s10865-007-9130-7] [Medline: 17899351]
33. 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]
34. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [doi: 10.1093/sleep/14.6.540] [Medline: 1798888]
35. Kendzerska TB, Smith PM, Brignardello-Petersen R, Leung RS, Tomlinson GA. Evaluation of the measurement properties of the Epworth sleepiness scale: a systematic review. Sleep Med Rev 2014 Aug;18(4):321-331. [doi: 10.1016/j.smrv.2013.08.002] [Medline: 24135493]
36. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993 Nov;4(5):353-365. [doi: 10.2165/00019053-199304050-00006] [Medline: 10146874]
37. DiBonaventura M, Richard L, Kumar M, Forsythe A, Flores NM, Moline M. The association between insomnia and insomnia treatment side effects on health status, work productivity, and healthcare resource use. PLoS One 2015 Oct 1;10(10):e0137117 [FREE Full text] [doi: 10.1371/journal.pone.0137117] [Medline: 26426805]
38. Selig JP, Preacher KJ. Mediation models for longitudinal data in developmental research. Res Human Dev 2009 Jun 01;6(2-3):144-164. [doi: 10.1080/15427600902911247]
39. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Soft 2012;48(2):1-36 [FREE Full text] [doi: 10.18637/jss.v048.i02]
40. Hayes AF. Beyond Baron and Kenny: statistical mediation analysis in the new millennium. Commun Monographs 2009 Dec;76(4):408-420. [doi: 10.1080/03637750903310360]
41. 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]
42. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
43. Palagini L, Moretto U, Dell'Osso L, Carney C. Sleep-related cognitive processes, arousal, and emotion dysregulation in insomnia disorder: the role of insomnia-specific rumination. Sleep Med 2017 Feb;30:97-104. [doi: 10.1016/j.sleep.2016.11.004] [Medline: 28215272]
44. Bonnet MH, Arand DL. Hyperarousal and insomnia: state of the science. Sleep Med Rev 2010 Feb;14(1):9-15. [doi: 10.1016/j.smrv.2009.05.002] [Medline: 19640748]
45. Mulgrew AT, Ryan CF, Fleetham JA, Cheema R, Fox N, Koehoorn M, et al. The impact of obstructive sleep apnea and daytime sleepiness on work limitation. Sleep Med 2007 Dec;9(1):42-53. [doi: 10.1016/j.sleep.2007.01.009] [Medline: 17825611]
46. Birrer D, Röthlin P, Morgan G. Mindfulness to enhance athletic performance: theoretical considerations and possible impact mechanisms. Mindfulness (N Y) 2012 May 5;3(3):235-246. [doi: 10.1007/s12671-012-0109-2]
47. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods 2001 Dec;6(4):330-351. [Medline: 11778676]

48. Graham JW, Cumsille PE, Elek-Fisk E. Methods for handling missing data. In: Schinka JA, Velicer WF, editors. Handbook of Psychology, Research Methods in Psychology. Volume 2. Hoboken, NJ, USA: John Wiley & Sons; 2003:87-114.

Abbreviations

ISI: Insomnia Severity Index

Edited by L Buis; submitted 27.06.22; peer-reviewed by C Liu; comments to author 25.07.22; revised version received 08.08.22; accepted 26.08.22; published 28.09.22.

Please cite as:
Espel-Huynh H, Baldwin M, Puzia M, Huberty J
The Indirect Effects of a Mindfulness Mobile App on Productivity Through Changes in Sleep Among Retail Employees: Secondary Analysis
JMIR Mhealth Uhealth 2022;10(9):e40500
URL: https://mhealth.jmir.org/2022/9/e40500
doi:10.2196/40500
PMID:36169994

Original Paper

Effect of Face-to-Face and WhatsApp Communication of a Theory-Based Health Education Intervention on Breastfeeding Self-Efficacy (SeBF Intervention): Cluster Randomized Controlled Field Trial

Farahana Mohamad Pilus[1*], DrPH; Norliza Ahmad[1*], PhD; Nor Afiah Mohd Zulkefli[1*], PhD; Nurul Husna Mohd Shukri[2*], PhD

[1]Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

[2]Department of Nutrition, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

[*]all authors contributed equally

Corresponding Author:
Norliza Ahmad, PhD
Department of Community Health
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
Jalan Hospital Serdang
Serdang, 43400
Malaysia
Phone: 60 97692424
Fax: 60 97692425
Email: lizaahmad@upm.edu.my

Abstract

Background: The exclusive breastfeeding rate in Malaysia is still not satisfactory. Previous studies have shown that breastfeeding self-efficacy is one of the determinants of exclusive breastfeeding, and it can be improved using social cognitive theory. WhatsApp, which is widely used among Malaysians, could be leveraged as a platform to deliver health education interventions.

Objective: This study aimed to develop, implement, and evaluate the effect of using a face-to-face and WhatsApp-based health education intervention based on social cognitive theory, namely the Self-Efficacy in Breastfeeding (SeBF) module, on mothers' self-efficacy, knowledge, and attitudes in a district in Selangor state.

Methods: This study was a 2-arm, parallel, single-blind, cluster randomized controlled field trial with an intervention and a control group involving primigravida or multigravida mothers who reside in a district in Selangor state and did not exclusively breastfeed during their previous pregnancy. All 12 maternity and pediatric clinics in this district were randomly divided into 6 intervention and 6 control groups. A total of 172 pregnant mothers were randomly assigned to the intervention group (n=86) or the control group (n=86). The control group received usual routine care. The primary outcome was breastfeeding self-efficacy, while secondary outcomes were knowledge and attitude toward breastfeeding. Each subject was assessed at 4 time points: at baseline, immediately after the intervention, 4 weeks post partum, and 8 weeks post partum. Generalized mixed model analysis was applied to measure the effect of health education on breastfeeding self-efficacy, knowledge, and attitude after the intervention.

Results: The response rate was 81% (139/172), with the dropout rate being 7% (6/86) in the intervention group and 31% (27/86) in the control group. In the intent-to-treat analysis, the intervention group showed a significant increase in the mean total breastfeeding self-efficacy score 8 weeks after delivery compared with the control group ($F_{21,601}$=111.73, P<.001). In addition, the mean total score for breastfeeding knowledge increased significantly in the intervention group after the intervention compared to the control group ($F_{21,601}$=8.33, P<.001). However, no significant difference was found in the mean total score for breastfeeding attitude after the intervention ($F_{21,602}$=5.50, P=.47).

Conclusions: Face-to-face and WhatsApp-based participation in the SeBF program, designed on the basis of social cognitive theory, contributed to improved self-efficacy and knowledge about breastfeeding. Further studies need to be conducted with a

longer duration (until 6 months post partum) to evaluate its effectiveness in increasing exclusive breastfeeding. Furthermore, new strategies in health education need to be developed to improve breastfeeding attitudes.

Trial Registration: Thaiclinicaltrials.org TCTR20200213004; https://www.thaiclinicaltrials.org/show/TCTR20200213004

(JMIR Mhealth Uhealth 2022;10(9):e31996) doi:10.2196/31996

KEYWORDS

self-efficacy; breastfeeding; intervention; social cognitive theory

Introduction

Background

The overall prevalence of exclusive breastfeeding in Malaysia was 47.1% (95% CI 43.13-51.18) [1], which is below the national target of 70% by 2025 [2]. Previous studies have shown that breastfeeding self-efficacy is one of the determinants of exclusive breastfeeding [3,4]. Breastfeeding self-efficacy refers to a woman's confidence in breastfeeding ability with an infant [3,5]. Several factors are associated with breastfeeding self-efficacy, including support and guidance, experience and stress, postpartum experiences, and social environment [5-7]. Social cognitive theory (SCT) has been shown to be effective in previous breastfeeding self-efficacy intervention studies [8-13]. SCT involves a cognitive dynamic process that assesses individuals' beliefs and ability to engage in healthy behaviors [8].

On the other hand, several methods have been used to deliver knowledge and skills to mothers, including face-to-face conversations, phone calls, and web-based applications [5,10-12,14,15]. WhatsApp, a cross-platform application that works on all major smartphone platforms such as Android, iPhone, and Windows Mobile, has approximately 500 million users worldwide [16]. According to a recent study [17], WhatsApp is preferred by 98.7% of Malaysian respondents, while Facebook Messenger is preferred by 54% of them. Given the widespread use of WhatsApp, we sought to evaluate the effectiveness of its use in providing health education to pregnant women.

Objectives

This study aimed to determine the effect of an SCT-based intervention called Self-Efficacy in Breastfeeding (SeBF) to improve breastfeeding self-efficacy through face-to-face communication and WhatsApp.

Methods

Study Design

This 2-arm, parallel, single-blind cluster randomized controlled field trial, comprising an intervention group and a control group, was conducted at maternal and child health clinics in Hulu Langat district, Selangor, Malaysia. The health clinics are considered a cluster for this study. The intervention group received the SeBF intervention, whereas the control group received standard routine brief counselling by health care personnel about breastfeeding and breastfeeding pamphlets.

Recruitment and Inclusion Criteria

Pregnant women between 34 and 37 weeks of gestation, who presented for antenatal care at maternal and child health clinics, were offered participation in this study, and their eligibility was assessed. Eligible mothers were primigravida or multigravida women who had not exclusively breastfed during a previous pregnancy and had a cell phone with Internet access and the WhatsApp app. Mothers who were taking medications such as anticancer drugs and those who had medical or pregnancy-related complications that hindered or complicated breastfeeding (heart disease, cancer, nephritis, active or untreated tuberculosis, HIV or AIDS, active breast herpes lesions, and severe malnutrition) were excluded from this study [14]. During recruitment, all participants have been informed that intervention was being offered. Mothers who expressed interest were provided with participant information sheets and informed consent forms.

Randomization and Allocation Concealment

The cluster comprised 12 maternal and child health clinics in Hulu Langat District. All chosen health clinics were randomly allocated into intervention and control groups by a nurse who was not involved in this study. All selected clinics were labeled, and Stat Trek software was used to perform simple randomization with a 1:1 allocation ratio [15]. The researcher was only aware of the intervention group's assignment after the randomization procedure was completed. Antenatal mothers who met the eligibility requirements were recruited and consented with an equal number of participants assigned to each clinic. Participants were blinded to group assignments. Participants were blinded to the fact that awareness of being part of the control group could influence their responses to the questionnaires.

Sample Size Calculation

The sample size was estimated on the basis of Lemeshow and Lwanga's Sample Size Determination in Health Studies [18]. The formula for 2 population proportions was used for hypothesis testing purposes. The sample size calculated was on the basis of a 30% increase in breastfeeding self-efficacy in the control group and a 55% increase in breastfeeding self-efficacy in the intervention group [11], with an α of .05 and β of .20, an intraclass correlation coefficient of 0.05 [19], attrition rate of 20% [20], and an average cluster size of 20 with a design effect of 2.45. After adjusting for the clustered design effect, the final minimum sample size required was 160 participants, with 80 participants in the intervention and control groups.

Intervention Module

The SeBF intervention was a newly developed module to improve breastfeeding self-efficacy among mothers. The development was based on SCT and prior intervention studies [5,8-10,12,15,18]. Breastfeeding self-efficacy was the main aim as it is one of the important determinants for successful exclusive breastfeeding [21]. Based on this intervention's success, it will be further evaluated for exclusive breastfeeding purposes.

This newly developed SeBF module was consulted and discussed with 2 Public Health Medicine Specialists and a Nutrition Specialist. This intervention module applied the SCT constructs such as observational learning, personal experience, verbal persuasion, problem-solving, self-efficacy, and outcome expectation. The SCT constructs used in the SeBF module showed in Multimedia Appendix 1.

The SeBF module was developed to be delivered face to face and through WhatsApp. The intervention consisted of training and reinforcement phases. The training phase involved a face-to-face session of 30 minutes, while the reinforcement phase involved WhatsApp messages weekly until 4 weeks post partum. The WhatsApp messages were sent every Monday at 2:30 PM for 15 minutes. The WhatsApp group function was used to distribute information, concerns, and issues; clarify any misunderstandings about breastfeeding practice; and provide a reminder to all participants.

The privacy and confidentiality of the participants were protected via a private group formation on WhatsApp. Thus, no other person apart from those recruited by the researcher could have access to the group. In addition, no personal information was exposed in the WhatsApp group. This module has been pilot-tested among 30 antenatal mothers not included in the main study. The SeBF intervention was delivered by a researcher who is also a medical doctor.

Study Instruments

The primary outcome of this intervention study was self-efficacy in breastfeeding scores, and the secondary outcome represents scores on knowledge and attitudes toward breastfeeding. These outcomes were assessed in the intervention and control groups immediately post intervention, 4 weeks post partum, and 8 weeks post partum.

Breastfeeding Self-efficacy

The 13-item Breastfeeding Self-Efficacy Scale–Short Form was used to assess mothers' confidence in their ability to successfully breastfeed their infant [22,23]. All items are preceded by the phrase “I can always” and are anchored on a 5-point Likert scale, where 1=“not at all confident” and 5=“always confident” [10]. Each item is presented positively, and the sum of the scores gives a range of 14 to 65, with higher scores indicating greater breastfeeding self-efficacy [10]. The Cronbach α coefficient was .90 [24].

Knowledge on Breastfeeding

A validated questionnaire with 47 items was used to assess knowledge about breastfeeding. These included general knowledge, benefits to mothers and babies, effective feeding method, duration of breastfeeding, expressed breast milk (EBM), storage of EBM, complementary foods, and problems with breastfeeding [25]. Each item had categorical responses of “True,” “False,” or “Not sure.” A correct answer was scored as “1,” while a wrong and an unsure answer were scored as “0.” The total knowledge score ranged from “0” to “47,” with higher scores representing more knowledge. The internal consistency of Cronbach α was .70 [25].

Breastfeeding Attitude

The 16-item Iowa Infant Feeding Attitude Scale was used to assess mothers' attitudes toward breastfeeding [26,27]. Mothers were asked to indicate the extent to which they agreed with each statement on a 5-point Likert scale ranging from 1=“strongly disagree” to 5=“strongly agree.” Items endorsing formula feeding were reverse-scored (1=5, 2=4, 3=3, 4=2, and 5=1), and an overall attitude score was calculated from the equally weighted sum of responses to each item. The total attitude scores ranged from “16” for a positive attitude toward formula breastfeeding to a maximum score of “80” for an attitude favoring breastfeeding. The Cronbach α for this instrument was ≥.85 [28].

Data Collection

The data collection for this study was conducted from January to December 2020. Baseline data were collected after participants' recruitment. A second assessment was conducted immediately after the training phase, followed by 4 and 8 weeks post partum. Concurrently, data from the control group were also collected at the same 4 time points. Attendance at the maternity and child health clinic was severely affected by the pandemic COVID-19. For the 4- and 8-week postpartum follow-ups, data collection was switched from hard copy self-administered questionnaires to Google Forms.

Statistical Analysis

SPSS (version 25; IBM Corp) was used for the analyses. Shapiro-Wilk, Komolgorov-Smirnov, and histogram tests were used to determine normal distribution. In the descriptive study, data were presented as mean, SD, frequency, and percentage values. The Pearson chi-square or Fisher exact test was used to determine the homogeneity of baseline data between the intervention and control groups for categorical data, and the Student *t* test was used for continuous data. The effectiveness of the intervention was determined using generalized mixed model analysis (GLMM), which controlled for baseline data covariates such as age, ethnicity, education level, maternal employment, and monthly family income. A significance level of .05 with a 95% CI was used for the study. Thus, the null hypothesis with a *P* value of >.05 was rejected.

Ethics Approval

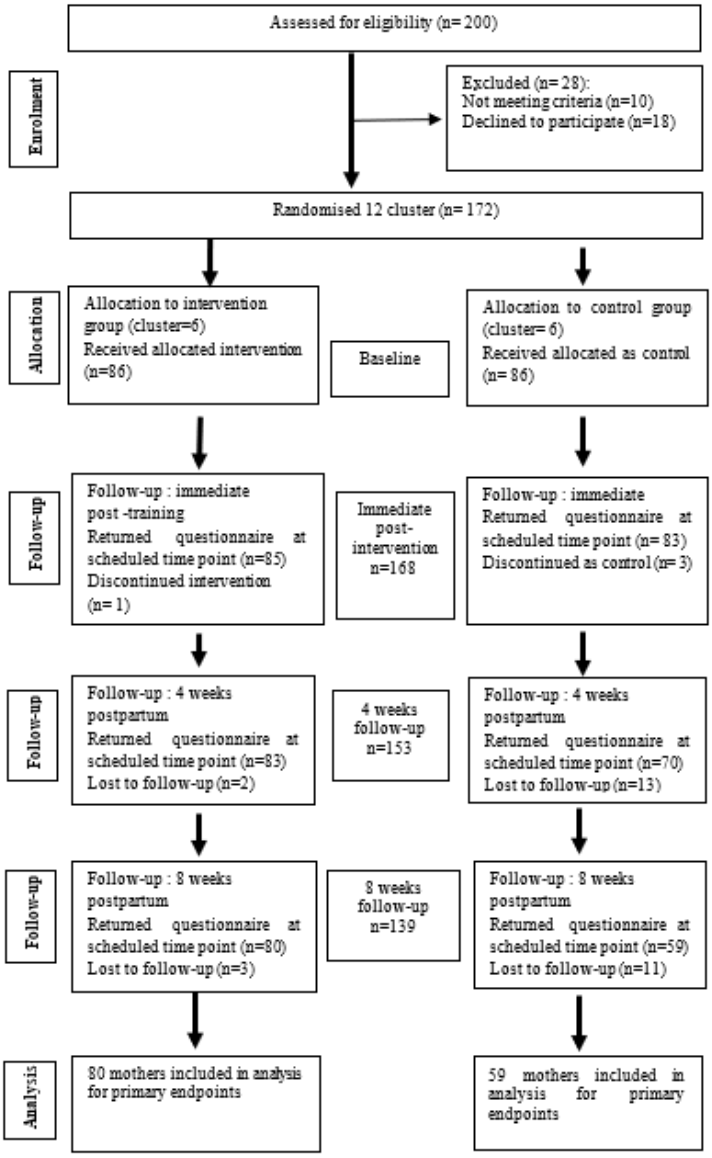
The National Medical Research Registry granted ethical approval for this study (NMRR-19-2712-50586). Each respondent provided written and informed consent during the data collection process. All participants’ information was kept strictly confidential. The study was prospectively registered in the Thai Clinical Trial Registry with identification number TCTR20200213004.

Results

Response Rate

A total of 139 mothers completed all 4 points of follow-up, resulting in a response rate of 81% (139/172) 8 weeks after delivery. The dropout rate was 31% (27/86) in the control group and 7% (6/86) in the intervention group. Figure 1 summarizes the final research schedule based on the CONSORT (Consolidated Standards of Reporting Trials) statement (Multimedia Appendix 2) [29].

Figure 1. Sample recruitment and dropouts throughout the study period.



Baseline Data

Table 1 provides an overview of the baseline data for the respondents. There were no significant differences in categorical or continuous variables between the intervention and control groups (P>.05). The household income in Malaysia was categorized as the B40 (below 40% of population) group, the M40 (middle 40% of population) group, and the T20 (top 20% of the population) group [30].

Table 1. Distribution of continuous and categorical variables at baseline (n=172).

Variables	Intervention (n=86)	Control (n=86)	Difference between the conditions	
			Fisher exact test or *t* test (*df*)	*P* value
Mother's current age (years), n (%)			N/A[a]	.25
19-24	24 (28)	18 (21)		
25-35	57 (66)	66 (77)		
36-45	5 (6)	2 (2)		
Ethnicity, n (%)			N/A	.09
Malay	73 (85)	75 (87)		
Chinese	4 (5)	2 (2)		
Indian	3 (3)	8 (9)		
Others	6 (7)	1 (1)		
Highest education level, n (%)			N/A	.48
Primary	2 (2)	5 (6)		
Secondary	28 (33)	25 (29)		
Higher education	56 (65)	56 (65)		
Mother's employment status, n (%)			N/A	.33
Unemployed	28 (32)	24 (28)		
Self-employed	4 (5)	9 (10)		
Government sector	7 (8)	11 (13)		
Private sector	47 (55)	42 (49)		
Total monthly household income (Malaysian Ringgit [RM]; 1 RM=US $0.22)[b], n (%)			N/A	.25
B40 (<RM 4850)	62 (72)	58 (67)		
M40 (RM 4850-10,959)	23 (27)	23 (27)		
T20 (>RM 10,960)	1 (1)	5 (6)		
Total score for breastfeeding self-efficacy, mean (SD)	22.44 (6.82)	24.38 (8.61)	–1.63 (169)	.11
Total score for breastfeeding knowledge, mean (SD)	22.44 (6.83)	24.12 (8.27)	–1.45 (169)	.15
Total score for breastfeeding attitudes, mean (SD)	56.27 (7.00)	56.97 (6.69)	–0.67 (169)	.51

[a]N/A: not applicable.

[b]Household income was categorized on the basis of the Department of Statistics Malaysia's classification of B40, M40, and T20.

Primary and Secondary Outcomes

Table 2 compares breastfeeding self-efficacy, knowledge about breastfeeding, and attitudes toward breastfeeding between the intervention and control groups at baseline, immediately after training, and 4 and 8 weeks post partum. Bivariate analyses show no significant difference between intervention and control groups for all time points, except for the total score for knowledge 8 weeks post partum.

Table 3 shows the GLMM results for the values of self-efficacy, knowledge, and attitude toward breastfeeding after controlling for covariates. The results show a significant difference between the intervention and control groups for self-efficacy and knowledge of breastfeeding ($F_{21,601}$=111.728, P<.001 and $F_{21,601}$=8.331, P<.001, respectively). However, there was no significant difference between the two groups for attitudes toward breastfeeding.

Figure 2 shows that breastfeeding self-efficacy improved in the intervention group over all 4 time points. The control group had an almost identical pattern to that of the intervention group. Although the control group showed an almost similar trend to the intervention group, it did not reach a high level of self-efficacy at 8 weeks after the intervention. Regarding breastfeeding knowledge, Figure 3 shows that the overall rating of breastfeeding knowledge improved for all respondents in the intervention group compared to that in the control group. Figure 4 shows that attitudes toward breastfeeding did not improve after the intervention study.

Table 2. Breastfeeding self-efficacy, knowledge, and attitudes between intervention and control groups at baseline, immediately after the intervention, and 4 and 8 weeks post partum.

Variables	At baseline	Immediately post intervention	4 weeks post partum	8 weeks post partum
Total score of self-efficacy				
Intervention group score, mean (SD)	22.44 (6.82)	24.02 (6.37)	52.60 (9.29)	53.34 (9.16)
Control group score, mean (SD)	24.38 (8.61)	24.37 (8.55)	51.22 (9.32)	51.89 (8.44)
t test (*df*)	–1.63 (169)	–0.30 (169)	0.95 (161)	0.90 (145)
P value	.11	.76	.34	.37
95% CI for difference of means	–6.43 to –0.71	–4.28 to 0.41	–1.50 to 4.26	–1.58 to 4.20
Total score of knowledge				
Intervention group score, mean (SD)	22.44 (6.83)	24.02 (6.405)	26.10 (5.66)	27.54 (5.98)
Control group score, mean (SD)	24.12 (8.27)	24.12 (8.217)	25.11 (7.43)	24.57 (7.41)
t test (*df*)	–1.45 (169)	–0.08 (169)	0.94 (157)	2.54 (141)
P value	.15	.94	.35	.01[a]
95% CI for difference of means	–3.97 to 0.61	–2.32 to 2.13	–1.09 to 3.08	0.65 to 5.29
Total score of attitudes				
Intervention group score, mean (SD)	56.27 (7.0)	56.26 (6.99)	56.71 (5.53)	57.35 (6.06)
Control group score, mean (SD)	56.97 (6.7)	56.98 (6.66)	57.91 (7.789)	58.88 (8.56)
t test (*df*)	–0.67 (170)	–0.69 (170)	–1.13 (158)	–1.25 (140)
P value	.50	.49	.27	.24
95% CI for difference of means	–2.76 to 1.37	–0.72 to 1.04	–3.33 to 0.93	–4.11 to 1.05

[a]Statistically significant at $P<.05$.

Table 3. The effect of health education on mothers' overall ratings of self-efficacy, knowledge, and attitudes toward breastfeeding.

Outcomes and parameters	*F* test (*df*)	*P* value[a]
Total scores of breastfeeding self-efficacy		
Group	0.85 (1, 601)	.36
Time	413.95 (3, 601)	<.001[b]
Group×time	111.73 (21, 601)	<.001[b]
Total scores of breastfeeding knowledge		
Group	6.38 (1, 601)	.02[b]
Time	4.29 (3, 601)	.005[b]
Group×time	8.33 (21, 601)	<.001[b]
Total scores of breastfeeding attitudes		
Group	0.91 (1, 602)	.34
Time	0.38 (3, 602)	.77
Group×time	5.50 (21, 602)	.47

[a]Using a generalized linear mixed model adjusted for respondents' age, ethnicity, level of education, employment, and household income.

[b]Statistically significant at $P<.05$.

Figure 2. Total scores on self-efficacy, showing the interaction between group and time, for all respondents.

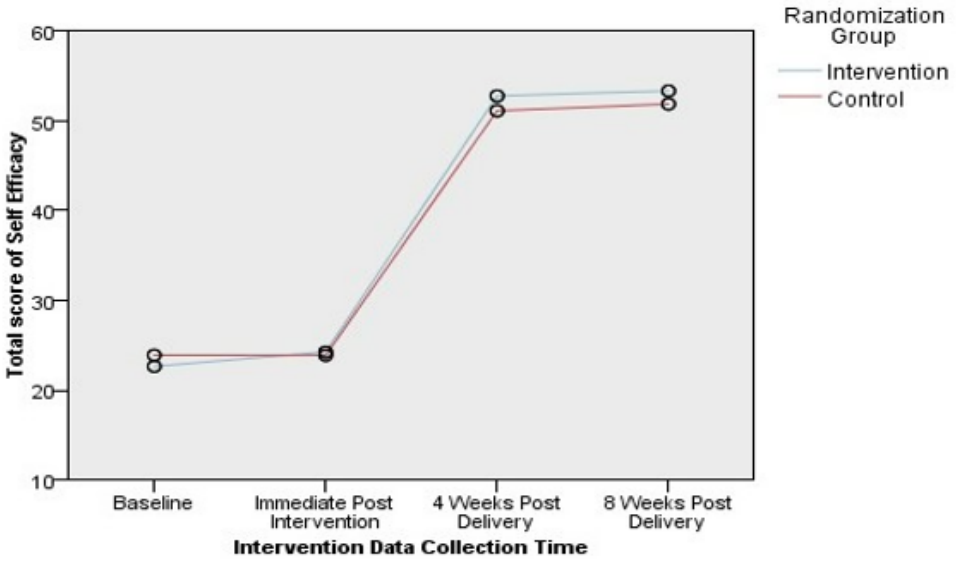


Figure 3. Total scores on breastfeeding knowledge, showing the interaction between group and time, for all respondents.

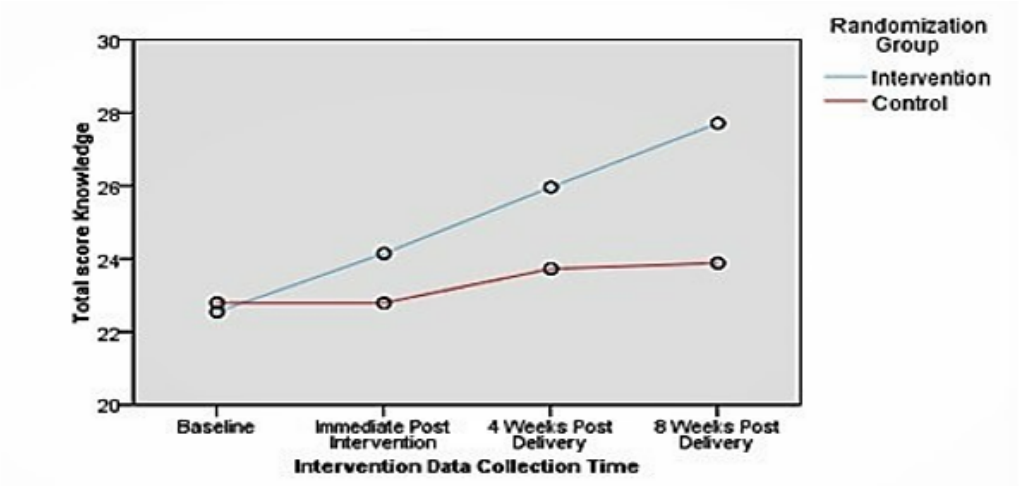
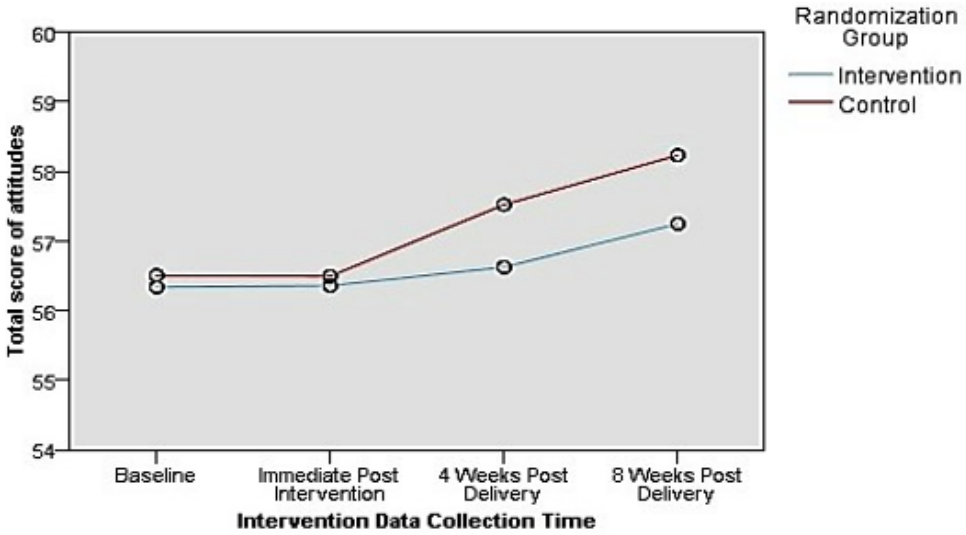


Figure 4. Total scores of attitudes on breastfeeding showing the interaction between group and time, for all respondents.



Discussion

Principal Findings

The purpose of the current study was to develop, implement, and evaluate the effect of the SeBF intervention on breastfeeding self-efficacy among antenatal mothers in Hulu Langat District, Selangor. The module was developed on the basis of SCT and delivered face to face and on WhatsApp. The findings showed that the intervention group had significantly increased their breastfeeding self-efficacy and knowledge scores compared to the control group. These significant changes could be contributed by the use of the SCT, namely observational learning, personal experience, verbal persuasion, and problem-solving to fulfil the end expectation and breastfeeding self-efficacy [31-33]. It included an antenatal period that began at the earliest 34 weeks of pregnancy and a postpartum period that lasted until 8 weeks post partum. It helped ensure that all intervention participants received an adequate and optimal dose of the SeBF intervention during the most critical period, from the third trimester of pregnancy to 8 weeks post partum [32]. Numerous approaches, including educational talk, practical breastfeeding videos, model demonstrations, and group discussions, contributed to respondents' increased self-efficacy. The usage of mobile technology, specifically WhatsApp, may have improved the primary and secondary outcomes. It facilitated the communication between the researcher and the participants in the intervention group. Hence, any problems faced by the mothers can be solved immediately. WhatsApp was chosen because it is a popular social media platform among Malaysians [30]. Therefore, it is readily used rather than requesting the respondents to install other new applications for this study.

Comparison With Prior Work

This study's findings are consistent with those of a previous study conducted in Iraq [13], indicating a significant difference in the change in mean breastfeeding self-efficacy scores between mothers who received the intervention and those who did not receive the intervention by applying SCT in their study. This study found that although all pregnant women in the study had some prenatal visits before enrollment, their breastfeeding knowledge was poor. This suggests the need to educate women about breastfeeding during their routine antenatal appointments [34]. Overall knowledge scores were higher in the intervention group than in the control group. An Iranian study found that women who received antenatal education have significantly higher knowledge scores, which resulted in higher mean breastfeeding self-efficacy scores; 53.98 (SD8.50) in the intervention group and 43.41 (SD 8.12) in the control group (P=.001) [13]. Another study in Canada reported the same result, with a significant increase in participants' knowledge scores following the educational breastfeeding intervention (mean knowledge scores of 24.14, SD 4.08 post intervention vs 11.43, SD 4.78 before the intervention, P=.001) [35].

There is no significant difference in the mean total score of breastfeeding attitude between intervention and control groups in this study. Our finding is consistent with a study in Jordan, which found that despite an improvement in respondents' self-efficacy for breastfeeding, their attitude ratings remained unchanged [27]. Furthermore, a study conducted in India found that the number of mothers with a favorable attitude dropped from the first to the third follow-up visit, despite improved knowledge at the baseline visit [36]. The most frequently cited reasons for their change in attitude toward breastfeeding were a perception of insufficient milk secretion, concern for the baby's health, concern that the baby would not gain enough weight on breastfeeding alone until 6 months, and complications such as breast engorgement and sore nipples during breastfeeding [36]. Increasing the intensity of WhatsApp communication may help improve the breastfeeding attitude and subsequently increase breastfeeding self-efficacy [11,12].

Strengths and Limitations

This study was a randomized controlled field trial with a reasonable participation rate and adequate follow-up despite the COVID-19 pandemic during data collection. Randomized assignment to the intervention and control groups makes the two comparable and increases validity [36]. It is the first study to employ innovative teaching and follow-up techniques for breastfeeding intervention using WhatsApp. Modification of educational materials and linking of responses to more accessible and appealing social media platforms are critical components of modern education [37].

However, this study has some limitations. Because of the self-completed questionnaire used in this study, social desirability may have been observed, with individuals responding positively being considered successful breastfeeding mothers. This study may also not be relevant to individuals who do not have access to a smartphone, as the follow-up of the intervention was conducted via WhatsApp. Furthermore, it is difficult to track whether the intervention group read and digested the material distributed on WhatsApp. In the future, the researcher should ask respondents random questions to determine if they understood the intervention material.

Replicating this intervention may be difficult with limited human resources since it will impose an additional demand on health care personnel. While some health care facilities may be able to provide educational talk, group discussions, and WhatsApp group follow-ups, others may find it time consuming and difficult. Before adopting a clinic-level intervention, considerations of appropriateness, time, and human resources are required.

Conclusions

Participation in the SeBF intervention, a face-to-face and WhatsApp-based intervention using SCT, significantly improved self-efficacy and knowledge about breastfeeding. Nevertheless, this study showed that respondents' attitudes did not improve. This study showed that WhatsApp could be a practical tool in complementing breastfeeding health education to mothers. The user-friendly application can deliver simple and easy-to-read health messages and facilitate communication between health care providers and mothers. In future, the study period should extend to 6 months to examine the module's capability in attaining the exclusive breastfeeding goals. Furthermore, new

strategies in health education need to improve breastfeeding attitudes.

Acknowledgments

We would like to express our gratitude to the Director-General of Health Malaysia for authorizing the publication of this paper. We would also like to thank the Hulu Langat District Health Office and all participating clinics, as well as the mothers who participated in this study, for their excellent cooperation and support, especially during data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Social Cognitive Theory Constructs Used in SeBF Module.
[DOC File , 43 KB - mhealth_v10i9e31996_app1.doc]

Multimedia Appendix 2
CONSORT eHEALTH Checklist (V1.6.2).
[PDF File (Adobe PDF File), 90 KB - mhealth_v10i9e31996_app2.pdf]

References

1. Institute for Public Health (IPH), National Institutes of Health, Ministry of Health Malaysia. National Health and Morbidity Survey (NHMS) 2016: Maternal and Child Health (volume 2): Findings. 2016. URL: http://iku.moh.gov.my/images/IKU/Document/REPORT/2016/NHMS2016ReportVolumeII-MaternalChildHealthFindingsv2.pdf [accessed 2022-08-11]
2. Shohaimi NM, Mazelan M, Ramanathan K, Meor Hazizi MS, Leong YN, Cheong XB, et al. Intention and practice on breastfeeding among pregnant mothers in Malaysia and factors associated with practice of exclusive breastfeeding: A cohort study. PLoS One 2022 Jan 7;17(1):e0262401 [FREE Full text] [doi: 10.1371/journal.pone.0262401] [Medline: 34995321]
3. Loke AY, Chan LS. Maternal breastfeeding self-efficacy and the breastfeeding behaviors of newborns in the practice of exclusive breastfeeding. J Obstet Gynecol Neonatal Nurs 2013 Nov;42(6):672-684. [doi: 10.1111/1552-6909.12250] [Medline: 24128113]
4. Brockway M, Benzies K, Hayden KA. Interventions to Improve Breastfeeding Self-Efficacy and Resultant Breastfeeding Rates: A Systematic Review and Meta-Analysis. J Hum Lact 2017 Aug 23;33(3):486-499. [doi: 10.1177/0890334417707957] [Medline: 28644764]
5. Shariat M, Abedinia N, Noorbala A. Breastfeeding Self-Efficacy as a Predictor of Exclusive Breastfeeding?: A Clinical Trial. Iran J Neonatol 2018;9(3):26-34. [doi: 10.22038/IJN.2018.24694.1316]
6. Tang K, Gerling K, Chen W, Geurts L. Information and Communication Systems to Tackle Barriers to Breastfeeding: Systematic Search and Review. J Med Internet Res 2019 Sep 27;21(9):e13947 [FREE Full text] [doi: 10.2196/13947] [Medline: 31573903]
7. Sikander S, Maselko J, Zafar S, Haq Z, Ahmad I, Ahmad M, et al. Cognitive-behavioral counseling for exclusive breastfeeding in rural pediatrics: a cluster RCT. Pediatrics 2015 Feb;135(2):e424-e431. [doi: 10.1542/peds.2014-1628] [Medline: 25583916]
8. Ansari S, Abedi P, Hasanpoor S, Bani S. The Effect of Interventional Program on Breastfeeding Self-Efficacy and Duration of Exclusive Breastfeeding in Pregnant Women in Ahvaz, Iran. Int Sch Res Notices 2014 Aug 19;2014:510793-510796 [FREE Full text] [doi: 10.1155/2014/510793] [Medline: 27379293]
9. Noel-Weiss J, Rupp A, Cragg B, Bassett V, Woodend AK. Randomized controlled trial to determine effects of prenatal breastfeeding workshop on maternal breastfeeding self-efficacy and breastfeeding duration. J Obstet Gynecol Neonatal Nurs 2006 Sep;35(5):616-624. [doi: 10.1111/j.1552-6909.2006.00077.x] [Medline: 16958717]
10. Wu DS, Hu J, McCoy TP, Efird JT. The effects of a breastfeeding self-efficacy intervention on short-term breastfeeding outcomes among primiparous mothers in Wuhan, China. J Adv Nurs 2014 Aug 08;70(8):1867-1879. [doi: 10.1111/jan.12349] [Medline: 24400967]
11. Javorski M, Rodrigues AJ, Dodt RCM, Almeida PCD, Leal LP, Ximenes LB. Effects of an educational technology on self-efficacy for breastfeeding and practice of exclusive breastfeeding. Rev Esc Enferm USP 2018 Jun 11;52:e03329 [FREE Full text] [doi: 10.1590/S1980-220X2017031803329] [Medline: 29898169]
12. Chan MY, Ip WY, Choi KC. The effect of a self-efficacy-based educational programme on maternal breast feeding self-efficacy, breast feeding duration and exclusive breast feeding rates: A longitudinal study. Midwifery 2016 May;36:92-98. [doi: 10.1016/j.midw.2016.03.003] [Medline: 27106949]
13. Piro SS, Ahmed HM. Impacts of antenatal nursing interventions on mothers' breastfeeding self-efficacy: an experimental study. BMC Pregnancy Childbirth 2020 Jan 06;20(1):19 [FREE Full text] [doi: 10.1186/s12884-019-2701-0] [Medline: 31906881]

14. Yadav H. BREASTFEEDING PRACTICES IN A RURAL COMMUNITY IN KEDAH. JUMMEC 2010 Jun 28;13(1):38-44. [doi: 10.22452/jummec.vol13no1.11]
15. Alford SH, Leadbetter S, Rodriguez JL, Hawkins NA, Scholl LE, Peipins LA. Cancer screening among a population-based sample of insured women. Prev Med Rep 2015;2:15-20 [FREE Full text] [doi: 10.1016/j.pmedr.2014.11.004] [Medline: 26844046]
16. Khanna V, Sambandam SN, Gul A, Mounasamy V. "WhatsApp"ening in orthopedic care: a concise report from a 300-bedded tertiary care teaching center. Eur J Orthop Surg Traumatol 2015 Jul;25(5):821-826. [doi: 10.1007/s00590-015-1600-y] [Medline: 25633127]
17. Müller J. Smartphone market in Malaysia - Statistics and facts. Statista. 2021. URL: https://www.statista.com/topics/6615/smartphones-in-malaysia/#dossierKeyfigures [accessed 2022-08-11]
18. Ogston SA, Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Adequacy of Sample Size in Health Studies. Biometrics 1991 Mar;47(1):347. [doi: 10.2307/2532527]
19. van Breukelen GJP, Candel MJJM. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient!. J Clin Epidemiol 2012 Nov;65(11):1212-1218 [FREE Full text] [doi: 10.1016/j.jclinepi.2012.06.002] [Medline: 23017638]
20. Bunik M, Shobe P, O'Connor ME, Beaty B, Langendoerfer S, Crane L, et al. Are 2 weeks of daily breastfeeding support insufficient to overcome the influences of formula? Acad Pediatr 2010 Jan;10(1):21-28. [doi: 10.1016/j.acap.2009.09.014] [Medline: 20129478]
21. Awaliyah SN, Rachmawati IN, Rahmah H. Breastfeeding self-efficacy as a dominant factor affecting maternal breastfeeding satisfaction. BMC Nurs 2019 Aug 16;18(Suppl 1):30 [FREE Full text] [doi: 10.1186/s12912-019-0359-6] [Medline: 31427892]
22. Dennis C. The breastfeeding self-efficacy scale: psychometric assessment of the short form. J Obstet Gynecol Neonatal Nurs 2003 Nov;32(6):734-744. [doi: 10.1177/0884217503258459] [Medline: 14649593]
23. Amini P, Omani-Samani R, Sepidarkish M, Almasi-Hashiani A, Hosseini M, Maroufizadeh S. The Breastfeeding Self-Efficacy Scale-Short Form (BSES-SF): a validation study in Iranian mothers. BMC Res Notes 2019 Sep 23;12(1):622 [FREE Full text] [doi: 10.1186/s13104-019-4656-7] [Medline: 31547846]
24. Harris-Luna ML, Badr LK. Pragmatic Trial to Evaluate the Effect of a Promotora Telephone Intervention on the Duration of Breastfeeding. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):738-748. [doi: 10.1016/j.jogn.2018.09.001] [Medline: 30292773]
25. Tengku Ismail TA, Sulaiman Z. Reliability and validity of a Malay-version questionnaire assessing knowledge of breastfeeding. Malays J Med Sci 2010 Jul;17(3):32-39 [FREE Full text] [Medline: 22135547]
26. Mora ADL, Russell DW, Dungy CI, Losch M, Dusdieker L. The Iowa Infant Feeding Attitude Scale: Analysis of Reliability and Validity1. J Appl Social Pyschol 1999 Nov;29(11):2362-2380. [doi: 10.1111/j.1559-1816.1999.tb00115.x]
27. Abuidhail J, Mrayan L, Jaradat D. Evaluating effects of prenatal web-based breastfeeding education for pregnant mothers in their third trimester of pregnancy: Prospective randomized control trial. Midwifery 2019 Feb;69:143-149. [doi: 10.1016/j.midw.2018.11.015] [Medline: 30513445]
28. Karande S, Perkar S. Do fathers′ attitudes support breastfeeding? a cross-sectional questionnaire-based study in Mumbai, India. Indian J Med Sci 2012;66(1):30. [doi: 10.4103/0019-5359.110861]
29. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345(sep04 1):e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
30. Department of Statistics Malaysia. 2020 Jul 10. URL: https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=TU00TmRhQ1N5TUxHVWN0T2VjbXJYZz09 [accessed 2022-08-11]
31. Bandura A. Social Cognitive Theory: An Agentic Perspective. Asian J Soc Psychol 1999;2(1):21-41. [doi: 10.1111/1467-839X.00024]
32. Dennis C. Theoretical underpinnings of breastfeeding confidence: a self-efficacy framework. J Hum Lact 1999 Sep 01;15(3):195-201. [doi: 10.1177/089033449901500303] [Medline: 10578797]
33. Edwards M, Jepson R, McInnes R. Breastfeeding initiation: An in-depth qualitative analysis of the perspectives of women and midwives using Social Cognitive Theory. Midwifery 2018 Feb;57:8-17. [doi: 10.1016/j.midw.2017.10.013] [Medline: 29175258]
34. M R, Shabadi N, Kulkarni P, Sunil Kumar D, Anup G, Narayana Murthy M. Effectiveness of educational intervention on breastfeeding among primi pregnant women- a longitudinal study. Clinical Epidemiology and Global Health 2020 Dec;8(4):1306-1311. [doi: 10.1016/j.cegh.2020.05.002]
35. Reyes C, Barakat-Haddad C, Barber W, Abbass-Dick J. Investigating the effectiveness of school-based breastfeeding education on breastfeeding knowledge, attitudes and intentions of adolescent females. Midwifery 2019 Mar;70:64-70. [doi: 10.1016/j.midw.2018.12.010] [Medline: 30580074]
36. Nilsson IMS, Kronborg H, Rahbek K, Strandberg-Larsen K. The significance of early breastfeeding experiences on breastfeeding self-efficacy one week postpartum. Matern Child Nutr 2020 Jul 05;16(3):e12986 [FREE Full text] [doi: 10.1111/mcn.12986] [Medline: 32543045]

37. 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

EBM: expressed breast milk
GLMM: generalized linear mixed model
SCT: social cognitive theory
SeBF: Self-Efficacy in Breastfeeding

Edited by L Buis; submitted 14.07.21; peer-reviewed by V Khanna; comments to author 27.01.22; revised version received 28.05.22; accepted 29.06.22; published 14.09.22.

Please cite as:
Mohamad Pilus F, Ahmad N, Mohd Zulkefli NA, Mohd Shukri NH
Effect of Face-to-Face and WhatsApp Communication of a Theory-Based Health Education Intervention on Breastfeeding Self-Efficacy (SeBF Intervention): Cluster Randomized Controlled Field Trial
JMIR Mhealth Uhealth 2022;10(9):e31996
URL: https://mhealth.jmir.org/2022/9/e31996
doi:10.2196/31996
PMID:36103244

Original Paper

Smartphone Apps for Patients With Hematologic Malignancies: Systematic Review and Evaluation of Content

Nerea Báez Gutiérrez[1*], PharmD; Héctor Rodríguez Ramallo[2*], PharmD; Marcos Fernández González[2], PharmD; Laila Abdel-Kader Martín[2], PhD

[1]Hospital Universitario Reina Sofía, Córdoba, Spain

[2]Hospital Universitario Virgen del Rocío, Sevilla, Spain

[*]these authors contributed equally

Corresponding Author:
Héctor Rodríguez Ramallo, PharmD
Hospital Universitario Virgen del Rocío
Avenida Manuel Siurot, s/n
Sevilla, 41013
Spain
Phone: 34 688 933 454
Email: hector.rodriguez.sspa@juntadeandalucia.es

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2022/10/e43412

Abstract

Background: Hematological malignancies (HMs) are a heterogeneous group of cancers representing a significant cause of morbidity and mortality. The chronification of HMs and the increasing use of smartphones may lead patients to seek their current unmet needs through mobile health apps.

Objective: The goal of this review was to identify and assess the quality of smartphone apps aimed at patients diagnosed with HMs.

Methods: A systematic search of apps that were aimed at patients diagnosed with HMs, accessed from a Spain IP address, and were available on the iOS (App Store) and Android (Google Play) platforms was conducted in November 2021. The search terms used were “hematology,” “blood cancer,” “leukemia,” “lymphoma,” and “myeloma” apps in English, Spanish, or both languages. The identified apps were downloaded and analyzed independently by 2 reviewers. Information about general app characteristics was collected. The Mobile Application Rating Scale (MARS) was used to assess quality. The resulting parameter of the analyses, the mean score of the apps, was compared by Student *t* test.

Results: Overall, 18 apps were identified; 7 were available on Android, 5 were available on iOS, and 6 were available on both platforms. All included apps were free; 3 were published in 2021, and among the apps published before 2021, only 6 were updated in 2021. Most (16/18, 89%) of the apps were aimed at patients with leukemia or lymphoma (16). The primary purposes of the apps were to provide general information about the condition (16/18, 89%) and monitor symptoms and clinical parameters (11/18, 61%). Health care professionals contributed to the development of 50% (9/18) of apps; 6 were owned and supported by scientific societies, and 3 were developed with the participation of health care professionals. The mean MARS score for the overall quality of the apps was 3.1 (SD 1.0). The engagement and aesthetics subscales were the lowest rated subscales, with only 44% (8/18) and 67% (12/18), respectively, of the apps obtaining acceptable scores. None of the included apps proved clinical efficacy through clinical trials in patients with HMs. Statistically significant differences were found in the MARS scores between operating systems (+1.0, *P*=.003) in favor of iOS apps. The participation of health care professionals in the development of the apps did not have a statistically significant impact on the MARS scores.

Conclusions: This systematic search and evaluation identified few acceptable quality mobile apps for patients with HMs. Current and future apps for patients with HMs should provide evidence-based valuable information, improve user engagement, incorporate functions according to patient preferences, and generate evidence regarding the efficacy of app use by patients with HMs.

(JMIR Mhealth Uhealth 2022;10(9):e35851) doi:10.2196/35851

KEYWORDS

hematological malignancies; mobile apps; smartphone; eHealth; mHealth; cancer; mobile app; mobile health; hematology

Introduction

Hematological malignancies (HMs) are a heterogeneous group of cancers that affect hematopoietic and lymphoid tissue. These disorders constitute 6.8% of all cancers, representing a significant cause of morbidity and mortality [1]. In the past decade, an increased incidence rate of HMs has been described [2,3], possibly due to a growth in life expectancy and the aging of the population [4].

HMs are frequently aggressive and require urgent, lengthy, and burdensome treatment. The complexity of HM treatments could reduce patient adherence and increase the risk of potential drug-drug interactions and adverse events [5-10]. Furthermore, patients with HMs and, specifically, patients with lymphoma and myeloma often experience psychological despair and poor quality of life throughout their illness course [11-13]. Although HM treatments add a burden to various aspects of a patient's life, the wide range of treatments now available has significantly improved the management of these patients, often transforming HMs into chronic diseases with long-term survival [14,15].

The chronification of HMs in the digital era may lead patients to seek their current unmet needs through internet-based health care, mobile health (mHealth), and, specifically, smartphone apps. An increase in mHealth use is a likely scenario, since the use of smartphones is growing, with estimates indicating that 77.6% of Europeans own a smartphone [16]. In this context, health apps designed for patients with HMs could serve as additional tools for telemedicine, patient education and life coaching, medication adherence, communication, and social media connections [17-19].

As the number of health-related apps has increased rapidly in the last decade, there is an increasing need for research to assess the quality of eHealth tools and identify patients' unmet needs.

This study aimed to identify apps that are accessible from Spain and are designed for patients diagnosed with HMs, analyze their characteristics, and evaluate their quality with the Mobile Application Rating Scale (MARS) [20].

Methods

Search Strategy

A systematic search of apps that were aimed at patients diagnosed with HMs, accessed from a Spain IP address, and were available on the iOS (App Store) and Android (Google Play) platforms was conducted from November 10, 2021, to November 25, 2021. The search was carried out following the PRISMA (Preferred Reporting Items for a Systematic Review and Meta-analysis) guidelines [21].

The search terms used were "hematology," "blood cancer," "leukemia," "lymphoma," "myeloma," and their equivalents in the Spanish language—"hematología," "cáncer de Sangre," "leucemia," "linfoma," and "mieloma," respectively.

The inclusion criteria used for selection were apps available in English or Spanish that were targeted at patients with HMs and included general information about HMs. The exclusion criteria used were apps that included inappropriate content (including horoscopes and astrology, among others), apps exclusively for fundraising for HMs that did not provide functions or tools designed for their use, or apps with nonfunctional links. The apps analyzed had to comply with all inclusion criteria to be included in the data extraction phase.

The names and descriptions of the apps from the search on Google Play and the App Store were selected against a priori selection criteria. The apps that met the inclusion criteria were downloaded for further screening using an Mi 9 Lite smartphone (Android version 10, Xiaomi Inc) for Google Play apps. For the apps downloaded from the App Store, an iPad Pro 2020 (Apple Inc) running iPadOS 15.1 was used. For apps duplicated across stores, the iOS app was evaluated because of the lower number of apps in the App Store compared with those in Google Play.

Data Extraction

Data were obtained from the App Store and Google Play online app descriptions (app characteristics and the narrative text) by 2 independent researchers. Data were extracted and entered into a structured Excel database (Microsoft Corp).

The variables collected for each app were the name, developer/owner name, type of developer/owner (commercial, scientific society, patient association), platform (Android or iOS), language (English or Spanish), app store category (eg, education, health and fitness, medicine), cost (€), date of publication (year), date of the last update (year), app file size (MB), app version, participation of health care professionals in the design and development of the app (yes/no), and HM targeted (eg, acute lymphocytic leukemia, multiple myeloma, Hodgkin lymphoma).

The participation of health care professionals was considered when specified in the app description or when the app was developed by a health organization (eg, a scientific society or a hospital). The purposes of the apps were classified into the following categories: assessment (eg, providing clinical scales, classifying adverse event severity, or interpreting laboratory findings), general information (eg, information about HMs, medications, or adverse events), the monitoring of clinical parameters (eg, register of laboratory parameters or symptoms), register of patient activities (eg, calendars for patients to add appointments and treatment administration), and contact with health care professionals or other patients.

A descriptive analysis was developed with continuous and discrete variables presented as mean and standard deviation and frequency and percentage, respectively. The means of

continuous variables were compared using a *t* test. Results with a *P* value of <.05 were considered statistically significant. The generated data were analyzed using SPSS (version 26.0, IBM Corp).

App Quality Evaluation

To evaluate app quality, the MARS tool was used [20]. The MARS is a validated system to assess health apps that has been described as the most comprehensive for evaluating technical information and capabilities of apps [20,22,23]. This tool has been widely used to evaluate health apps designed for many diseases [24-30].

The MARS comprises 23 items across 5 subscales:

- Engagement: evaluates the entertainment, interest, customization, interactivity, and adequacy of the target group
- Functionality: assesses the performance, ease of use, navigation, and gestural design of the app
- Aesthetics: examines the layout, graphics, and visual appeal of the app
- Information: assesses the accuracy of the description, establishment of goals, quality and quantity of information, visual information, credibility of the source, and evidence-based development of the app
- Subjective quality: determines willingness to recommend app, times app will be used, willingness to pay, and overall rating of app

Each criterion is evaluated from 1 to 5 (1=inadequate; 2=poor; 3=acceptable; 4=good; 5=excellent). A mean score of the 5 subscales is calculated to describe overall quality.

Before the app search was carried out, reviewers read and became acquainted with the MARS tool. All authors then discussed each rating criteria to achieve a consensus on how to apply them. The first app included was evaluated concurrently by all reviewers to ensure a common understanding and application of the MARS tool. The apps included were independently assessed by 2 reviewers, whose scores were then compared, and a final MARS score was agreed upon. A third reviewer was invited to resolve discrepancies if a consensus on the final MARS score was not reached.

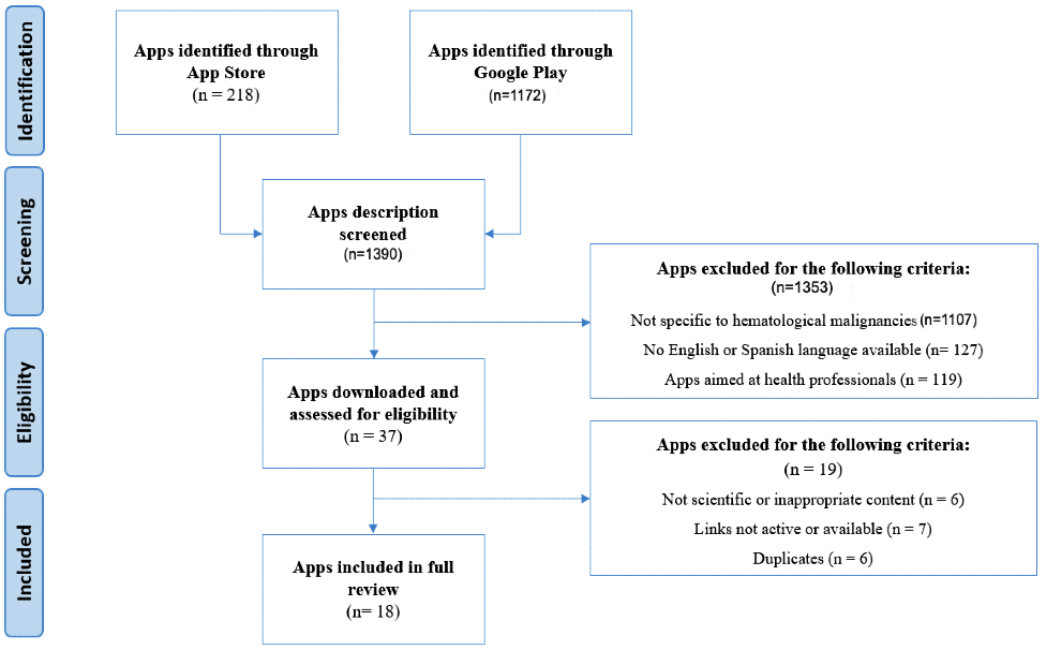
In order to quantify the subjectivity degree of the reviewers' evaluations, the interrater reliability of the quality scores between the independent evaluations was calculated using the intraclass correlation coefficient (ICC) for a 2-way random effects model.

Results

Overview

The titles, descriptions, and screenshots of 1390 apps were screened in a pair-review fashion. Overall, 18 apps met the inclusion criteria for a comprehensive evaluation with the MARS tool (Figure 1).

Figure 1. Prisma Flow-Chart.



App Characteristics

Table 1 displays the general features of the apps. Regarding the apps available on the platforms searched, 7 were available on Google Play (Android), 5 were available on the App Store (iOS), and 6 were available on both platforms. The apps were categorized in different sections in their respective platforms, as follows: medicine (11), health and well-being (4), education (1), news (1), and books and reference works (1). All apps evaluated were free. The average file size of the apps was 27.2 (SD 25.3) MB.

Table 2 displays apps' purposes and health care professional involvement in their design. All 7 apps with a unique purpose were designed to provide information about HMs and their management. Eleven apps had multiple purposes, the most frequent being providing information and monitoring symptoms and clinical parameters. Overall, 9 of the apps were developed

with the participation of health care professionals, 6 were owned and supported by scientific societies, and 3 were developed with the participation of physicians or nurses. None of the included apps reported clinical efficacy through clinical trials in patients with HMs.

Table 1. Smartphone app technical characteristics.

Name	Year of publication	Year of the last update	Platform		Language	
			Android	iOS	English	Spanish
ALL Manager	2019	2021		✓	✓	
CLL Manager	2017	2021		✓	✓	
CLL Watch and Wait Tracker	2017	2020	✓	✓	✓	
CML Life	2018	2019	✓	✓	✓	✓
CML Today	2015	2015	✓	✓	✓	✓
Don’t Walk Alone	2017	2018		✓	✓	
Focus on Lymphoma	2013	2018	✓		✓	
Hodgkin Lymphoma Manager	2019	2021		✓	✓	
Leucemia: Síntomas Y Tratamiento: FAQ	2018	2018	✓		✓	
Leukaemia (Leukaemia) News	2020	2020	✓		✓	
Leukemia: Causes, Diagnosis, and Treatment	2021	2021	✓		✓	
Leukemia Disease	2017	2017	✓		✓	✓
Leukemia Drug Tracker	2021	2021	✓		✓	
LLS CAR T	2020	2020	✓	✓	✓	
LLS Health Manager	2019	2021	✓	✓	✓	
LRF Understanding Lymphoma	2020	2021	✓	✓	✓	
Mieloma	2021	2021	✓			✓
Multiple Myeloma Manager	2015	2021		✓	✓	

Table 2. App purpose and health care professional involvement.

Name	Hematologic malignancy	Purpose					HPP[a]
		I[b]	M[c]	R[d]	C[e,f]	CP[g]	
ALL Manager	ALL[h]	✓	✓	✓	✓	✓	
CLL Manager	CLL[i]	✓	✓	✓	✓	✓	
CLL Watch and Wait Tracker	CLL	✓	✓				✓
CML Life	CML[j]	✓	✓				✓
CML Today	CML		✓	✓			✓
Don't Walk Alone	CLL	✓	✓		✓	✓	✓
Focus on Lymphoma	Lymphoma	✓	✓	✓		✓	✓
Hodgkin Lymphoma Manager	HL[k]	✓	✓	✓	✓	✓	
Leucemia: Síntomas Y Tratamiento: FAQ	Leukemia	✓					
Leukaemia (Leukaemia) News	Leukemia	✓					
Leukemia: Causes, Diagnosis, and Treatment	Leukemia	✓					
Leukemia Disease	Leukemia	✓					
Leukemia Drug Tracker	Leukemia	✓	✓	✓			
LLS CAR T	HM[l]	✓					✓
LLS Health Manager	HM	✓	✓	✓	✓	✓	✓
LRF Understanding Lymphoma	Lymphoma	✓					✓
Mieloma	MM[m]	✓					✓
Multiple Myeloma Manager	MM	✓	✓	✓	✓	✓	

[a]HPP: health care professional participation.
[b]I: general information.
[c]M: monitoring of symptoms and clinical parameters.
[d]R: register of patient activities.
[e]C: contact with health care professionals.
[f]This feature was included in the app but was not available for Spanish patients.
[g]CP: contact with other patients.
[h]ALL: acute lymphocytic leukemia.
[i]CLL: chronic lymphocytic leukemia.
[j]CML: chronic myeloid leukemia.
[k]HL: Hodgkin lymphoma.
[l]HM: hematological malignancy.
[m]MM: multiple myeloma.

App Quality Assessment

Tables 3 and 4 display the MARS subscales scores for the individual apps. The scores in the functionality section were similar between the apps. The movements between menus and screens were fast and satisfactory for most apps. The most remarkable differences were found in the information and aesthetics sections, for which only 67% (12/18) and 61% (11/18) of the apps showed acceptable scores (>3), respectively. None of the apps were tested in clinical trials.

The mean MARS score was 3.1 (SD 1.0), considering this value as acceptable. Table 5 contains the overall quality scores. For 5 apps, the mean MARS score was >4, meaning these apps were of good quality. On the other hand, 7 apps received a mean score of <3 and were classified as poor-quality apps.

iOS apps obtained better MARS scores when compared to apps for Android (+1.0) points when comparing apps based on their platforms. This difference was statistically significant for the mean MARS score (P=.003) and for all 5 domains: engagement (+1.0; P=.001), functionality (+0.6, P=.05), aesthetics (+1.1, P=.01), information (+1.0, P=.01), and subjective evaluation (+1.1, P=.002). For apps available on both platforms, the MARS mean scores were higher than those for apps only available on one of the platforms (+0.6, P=.19). The aesthetic domain score was statistically superior for apps available on both platforms (+1.2, P=.04). The participation of health care professionals

had a positive impact on the appraisals of the apps (+0.6 points), but no statistically significant differences were found (P=.22).

For the overall MARS ratings, the ICC was 0.94 (95% CI 0.75-0.98), confirming excellent interrater reliability. For the engagement, functionality, aesthetics, information, and subjective domains, the ICCs were 0.84, 0.86, 0.94, 0.95, and 0.93, respectively.

Table 3. The Mobile Application Rating Scale engagement and functionality subscales.

Name	Engagement, score					Functionality, score			
	Ent[a]	Int[b]	Cus[c]	Iy[d]	Tg[e]	Per[f]	EU[g]	Nav[h]	GD[i]
ALL Manager	4	4	4	4	4	4	4	4	4
CLL Manager	4	4	4	4	4	4	4	4	4
CLL Watch and Wait Tracker	2	3	3	2	4	5	3	4	4
CML Life	4	4	4	2	4	3	3	4	4
CML Today	2	3	3	2	4	5	4	4	4
Don’t Walk Alone	2	3	3	3	4	3	3	3	3
Focus on Lymphoma	4	4	4	3	5	4	5	4	5
Hodgkin Lymphoma Manager	4	4	4	4	4	4	4	4	4
Leucemia: Síntomas Y Tratamiento: FAQ	2	2	1	1	3	3	3	3	3
Leukaemia (Leukaemia) News	2	2	1	1	3	2	4	2	2
Leukemia: Causes, Diagnosis, and Treatment	2	1	1	1	1	2	2	1	1
Leukemia Disease	2	2	1	1	3	2	3	3	3
Leukemia Drug Tracker	2	1	2	2	3	2	3	3	3
LLS CAR T	4	4	1	1	4	5	4	5	5
LLS Health Manager	4	3	3	4	4	4	4	4	4
LRF Understanding Lymphoma	4	4	1	1	4	4	3	4	3
Mieloma	1	2	2	3	2	3	2	4	3
Multiple Myeloma Manager	4	4	4	4	4	4	4	4	4
Overall, mean (SD)	2.9 (1.1)	3 (1.1)	2.6 (1.3)	2.4 (1.2)	3.6 (0.9)	3.5 (1.0)	3.4 (0.8)	3.6 (0.9)	3.5 (1.1)

[a]Ent: entertainment.

[b]Int: interest.

[c]Cus: customization.

[d]Iy: interactivity.

[e]Tg: target group.

[f]Per: performance.

[g]EU: ease of use.

[h]Nav: navigation.

[i]GD: gestural design.

Table 4. The Mobile Application Rating Scale aesthetics and information subscales.

Name	Aesthetics, score			Information, score					
	Lay[a]	Gra[b]	VA[c]	AAD[d]	Goa[e]	QI[f]	QyI[g]	VI[h]	Cre[i]
ALL Manager	4	4	4	5	3	5	5	4	3
CLL Manager	4	4	4	5	3	5	5	4	3
CLL Watch and Wait Tracker	4	3	4	5	N/A[j]	4	3	N/A	3
CML Life	4	3	4	4	N/A	3	3	4	3
CML Today	4	3	3	4	4	N/A	N/A	4	3
Don’t Walk Alone	5	4	4	3	N/A	2	4	2	2
Focus on Lymphoma	5	4	4	5	N/A	5	5	4	3
Hodgkin Lymphoma Manager	4	4	4	5	3	5	5	4	3
Leucemia: Síntomas Y Tratamiento: FAQ	1	2	1	1	N/A	2	1	N/A	1
Leukaemia (Leukaemia) News	1	2	1	4	N/A	1	3	N/A	2
Leukemia: Causes, Diagnosis, and Treatment	1	2	1	2	N/A	1	2	N/A	1
Leukemia Disease	1	2	1	1	N/A	2	2	N/A	1
Leukemia Drug Tracker	2	1	2	4	N/A	3	1	N/A	1
LLS CAR T	5	5	4	5	N/A	4	4	5	3
LLS Health Manager	4	3	4	5	4	4	3	4	3
LRF Understanding Lymphoma	4	5	5	5	N/A	5	4	4	3
Mieloma	3	1	2	4	N/A	3	3	N/A	3
Multiple Myeloma Manager	4	4	4	5	3	5	5	4	3
Overall, mean (SD)	3.3 (1.5)	3.1 (1.2)	3.1 (1.4)	4 (1.4)	3.3 (0.5)	3.5 (1.5)	3.4 (1.4)	4 (0.7)	2.4 (0.8)

[a]Lay: layouts.

[b]Gra: graphics.

[c]VA: visual appeal.

[d]AAD: accuracy of app description.

[e]Goa: goals.

[f]QI: quality of information.

[g]QyI: quantity of information.

[h]VI: visual information.

[i]Cre: credibility.

[j]N/A: not applicable.

Table 5. The Mobile Application Rating Scale overall quality scores.

Name	Engagement (score), mean (SD)	Functionality (score), mean (SD)	Aesthetics (score), mean (SD)	Information (score), mean (SD)	Subjective quality (score), mean (SD)	Overall score, mean (SD)
ALL Manager	4 (0.0)	4 (0.2)	4 (0.0)	4.2 (1.0)	4.3 (0.8)	4.1 (0.1)
CLL Manager	4 (0.0)	4 (0.2)	4 (0.0)	4.2 (1.0)	4.3 (0.8)	4.1 (0.1)
CLL Watch and Wait Tracker	2.8 (0.8)	4 (0.7)	3.7 (0.5)	3.8 (1.0)	2.8 (0.4)	3.4 (0.6)
CML Life	3.6 (0.9)	3.5 (0.5)	3.7 (0.5)	3.5 (0.6)	2.4 (0.8)	3.3 (0.5)
CML Today	2.8 (0.8)	4.3 (0.4)	3.3 (0.5)	3.8 (0.5)	2.5 (0.9)	3.3 (0.7)
Don’t Walk Alone	3 (0.7)	3 (0.0)	2.8 (0.2)	2.6 (0.9)	1.8 (0.4)	2.7 (0.5)
Focus on Lymphoma	4 (0.7)	4.6 (0.5)	4.3 (0.5)	4.4 (0.9)	3.6 (0.5)	4.2 (0.4)
Hodgkin Lymphoma Manager	4 (0.0)	4 (0.2)	4 (0.0)	4.2 (1.0)	4.3 (0.8)	4.1 (0.1)
Leucemia: Síntomas Y Tratamiento: FAQ	1.8 (0.8)	3 (0.0)	1.3 (0.5)	1.3 (0.5)	1.5 (0.4)	1.8 (0.7)
Leukaemia (Leukaemia) News	1.8 (0.8)	2.5 (0.9)	1.3 (0.5)	2.5 (1.3)	1.8 (0.4)	2.0 (0.5)
Leukemia: Causes, Diagnosis, and Treatment	1.2 (0.5)	1.5 (0.5)	1.3 (0.5)	1.5 (0.6)	1.1 (0.2)	1.3 (0.2)
Leukemia Disease	1.8 (0.8)	2.8 (0.4)	1.3 (0.5)	1.5 (0.6)	1.3 (0.3)	1.8 (0.6)
Leukemia Drug Tracker	2 (0.7)	2.8 (0.4)	1.7 (0.5)	2.3 (1.5)	1.5 (0.4)	2.0 (0.5)
LLS CAR T	2.8 (1.7)	4.8 (0.4)	4.7 (0.5)	4.2 (0.8)	2.3 (0.8)	3.8 (1.1)
LLS Health Manager	3.6 (0.5)	4 (0.0)	3.7 (0.5)	3.8 (0.8)	2.5 (0.4)	3.5 (0.6)
LRF Understanding Lymphoma	2.8 (1.7)	3.5 (0.5)	4.7 (0.5)	4.2 (0.8)	3.1 (0.7)	3.7 (0.8)
Mieloma	2.0 (0.7)	3 (0.7)	2 (0.8)	3.3 (0.5)	1.7 (0.4)	2.4 (0.7)
Multiple Myeloma Manager	4 (0.0)	4 (0.2)	4 (0.0)	4.2 (1.0)	4.3 (0.8)	4.1 (0.1)
Overall	2.9 (0.9)	3.5 (0.8)	3.1 (1.3)	3.3 (1.1)	2.6 (1.1)	3.1 (1.0)

Discussion

Principal Findings

Health apps have the potential to become a standard of care for chronic patients. Apps have been associated with positive patient outcomes such as improved adherence to medication and quality of life and decreased use of health care–related resources [31]. Therefore, health apps may offer valuable tools for patients with HMs in a health care context [17,32-34].

This systematic review highlighted the limited number of apps available for patients with HMs. The number of apps identified in this study was significantly lower than previously published reviews on COVID-19, neoplasms, and specific types of cancer such as breast cancer [35-37]. The apps were developed for a narrow range of diseases that HMs encompass. Currently, patients with acute myeloid leukemia or non-Hodgkin lymphoma do not have access to eHealth tools in the main app stores that specifically target these diseases. In addition, there is a lack of apps specifically designed for children, regardless of them being a significant population of patients with HMs who have distinct needs [38,39].

The low number of apps found in this review may be due to several reasons. First, HMs are relatively rare; according to the Global Cancer Observatory, HMs account for an estimated 6.8% of all cancers worldwide [1]. Similarly, systematic searches for diseases with low prevalence tend to find few apps directed to patients [1,24,40,41]. Second, strict exclusion criteria were applied for app selection—specifically, the exclusion of apps aimed at health care professionals, which may encompass half of the health apps [42]. Third, HMs often affect older patients, an age group for which developers may often not design eHealth tools [43,44]. Older age has been associated with a lack of interest in eHealth and smartphones [45,46]. This population also faces additional barriers to using eHealth, such as physical disability and technology inexperience [47,48].

All apps included for evaluation were designed to serve as an information source or monitoring tool for medications and symptoms. The information was not frequently reviewed and updated, as only half of the HM apps (9/18, 50%) were updated during 2021. The management of HMs is in continuous change, often requiring guidelines and recommendations to be updated with frequency, often yearly. As an illustration of this constant evolution, the European Medicines Agency has recommended for approval an average of 5 drugs for HMs per year [49]. The growing arsenal of new treatments highlights the importance of regularly updating apps according to changes in guidelines and recommendations, as the management of HMs is a complex field of constant change [50].

Among the few apps that complied with the inclusion criteria for this study, the overall quality was acceptable. The highest rated subscale on the MARS was functionality due to app efficiency and ease of use. On the other hand, the engagement

subscale was the worst evaluated, specifically the customization and interactivity items. Poor results in the engagement domain could mean that even if quality apps include valuable tools, patients will use them for limited periods. Furthermore, the customization of app interactions by setting personalized reminders is perceived as highly beneficial by patients with HMs [51-53]. Finally, the credibility item was poorly evaluated, as the existence of commercial interests or the source's legitimacy could not be verified.

Currently, there is a lack of high-quality evidence on the eHealth needs of patients with HMs. Surveys about communication technologies conducted on hematology and oncology patients found that mobile phones were the most frequently used device to search for health-related information. Patients were highly interested in staying informed about health issues, disease prevention, healthy lifestyles, and general information about the disease [31,54]. Patients seem to be interested in eHealth helping with practical issues, such as appointment management, the provision of advice on disease and symptom management, and direct communication with health care professionals in addition to providing information. Patients were less interested in features that can add additional burdens to their daily activities, such as recording and monitoring medication, symptoms, and adverse events [31]. Patients with HMs also desire the creation of communication channels that allow health care professionals to answer their concerns rapidly [31]. A scoping review of virtual care in patients with HMs described high patient satisfaction with telemedicine interventions that allowed clinicians and nurses to communicate via phone calls or videoconferencing [33].

The 18 apps that were designed for patients with HMs and were evaluated in this review did not fulfill the previously described preferences. Although 17 apps provided information about the disease, most of the information did not focus on disease prevention, symptom management, or healthy lifestyles. In addition, direct communication with health care professionals for Spanish patients was not possible, and the management of medical appointments was not an included feature. Further research on the virtual care needs of patients with HMs is warranted; surveys and validated questionnaires explicitly designed for patients with HMs are needed to gain knowledge of the unmet needs of these patients. Understanding patient preferences through a continuous user-centered design is essential for the success of mHealth for patients with HMs [50,55]. The differences found between features included in apps and the preferences of patients with HMs highlight the apparent necessity of their participation in the design of health-related apps. The inclusion of health care professionals may not be sufficient to adequately provide quality apps that cover patient needs, as there is a divide between the expectations and needs of patients and those identified by health care professionals [56,57].

Among the stakeholders involved in the management of HMs, the participation of health care professionals in the development of mHealth apps is frequently described as a critical factor related to their general quality and, specifically, the quality of the information provided [24,25,58]. A systematic review that evaluated expert participation and adherence to medical evidence in mHealth apps showed that the lack of participation of health care professionals and inclusion of evidence-based information were persistent characteristics of the available health apps [58]. However, even when these factors are potentially related, expert involvement in the design does not guarantee adherence to evidence-based content [58,59]. In this review, half of the apps (9/18, 50%) targeting patients with HMs included the participation of health care professionals in the development phase. Nevertheless, we did not find statistically significant differences in the quality scores based on this participation. Furthermore, a cross-sectional survey of oncology and hematology patients found that more than 80% of patients would use an app if a health professional recommended it [60]. These data suggest that health care professionals may also play a crucial role in high-quality apps reaching patients with HMs.

Due to the complex nature of HMs, which often require highly complex and burdensome therapeutic protocols, we could presume that patients would benefit from additional tools that help manage different aspects of the disease. However, the role of apps in the care of patients with HMs is uncertain; a review exploring the research on eHealth intervention in HMs found that few trials were designed exclusively for patients with HMs. These studies generally included mixed cancer groups composed of a small sample of hematology patients; moreover, the principal study results were not presented separately [34]. Similarly, among the 18 apps evaluated in this review, none assessed the clinical impact of the intervention through clinical trials; this finding is common among studies that evaluate health apps, with authors often stating the need for studies that assess the efficacy of health apps [23-25,37,61,62]. Additional randomized controlled trials testing the efficacy of apps are warranted before implementing this technology for managing HMs, as eHealth may also increase the burden of treatment on vulnerable patients [63,64].

Recommendations and Future Research for HM App Development

Current and future apps for patients with HMs should focus on improving the lowest rated subscales of the MARS scale: engagement and aesthetics. Specifically, improvements are needed in the interactivity, customization, and entertainment features, as well as in graphical design and visual appeal. Further refinement of the information provided may be needed in order to deliver updated evidence-based information adapted to the necessities and preferences of patients with HMs. In this context, HM apps should be developed through a user-centered, collaborative design process with the participation of all stakeholders involved. The participation of health care professionals can increase the credibility of apps and trustworthiness of sources [24,25,58]. Furthermore, the participation of patients with HMs is essential for developing apps that fulfill their unmet needs, increasing their usability [56,57]. App developers and stakeholders should work to include apps in specialized and primary care assistance circuits by adding functions that facilitate the access of patients with HMs to health assistance through the management of appointments and direct communication with health care professionals [31,54]. Regular updates of app contents should be performed to

incorporate information that is consistent with recently generated evidence.

Future research in the field of HM apps should focus on gaining knowledge regarding the eHealth needs of patients with HMs, conducting surveys that investigate the changing needs of specific groups of patients with HMs, measuring the existence of health app safety concerns, and generating evidence-based data on the effectiveness of eHealth interventions through apps in patients with HMs.

Limitations

This systematic review has several limitations. First, the reviewers assigned a score for each category using brief descriptions of what makes an excellent, regular, or mediocre app. Thus, there was a degree of subjectivity when evaluating apps using the MARS [21,22]. However, the multiple reviewer design of this study may narrow the impact of this limitation because of the high degree of interreviewer agreement observed. Second, despite the MARS being a validated tool that is widely used to assess the quality of apps for many diseases [23-29], it does not consider aspects such as privacy, security, and the frequency of updates to apps, which have been considered aspects of particular relevance when evaluating the quality of health software [65]. Third, as health care professionals, the authors evaluated apps aimed at patients; this has limitations, as we may not fully comprehend their needs and preferences [56,57]. Fourth, the search was limited to 2 app platforms: Google Play and the App Store. Although these platforms encompass the vast majority of apps available on the market [66], several platforms (Windows Phone, Blackberry Market, etc) were not included in this review due to the lack of availability of devices for performing the search among the reviewers. Therefore, the possibility exists that some apps dedicated to HMs were missed.

Conclusion

The potential for eHealth to improve the care of patients with HMs exists. However, current apps need further refinement in order to provide evidence-based valuable information, keep users engaged, and provide a visually appealing interface. Moreover, there is a need to generate evidence on the efficacy of eHealth interventions in patients with HMs.

Future eHealth tools for patients with HMs should consider including all stakeholders in the design phase. Notably, the participation of patients in this phase will serve to design and implement tools according to their currently unmet needs.

Acknowledgments

The authors are immensely grateful to Dominica Melgar Carrascal for her valuable ideas and insight into hematological malignancy app potential and limitations from the perspective of patients with hematological malignancies.

Authors' Contributions

NBG and HRR conceived the study. NBG, HRR, and MFG collected the data and wrote the manuscript. NBG, HRR, MFG, and LAKM contributed to the interpretation of the results and reviewed the paper. All authors have reviewed and approved the last version of the manuscript.

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 Feb 04:209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. JAMA Oncol 2018 Nov 01;4(11):1553-1568 [FREE Full text] [doi: 10.1001/jamaoncol.2018.2706] [Medline: 29860482]
3. Cowan AJ, Allen C, Barac A, Basaleem H, Bensenor I, Curado MP, et al. Global burden of multiple myeloma: a systematic analysis for the Global Burden of Disease Study 2016. JAMA Oncol 2018 Sep 01;4(9):1221-1227 [FREE Full text] [doi: 10.1001/jamaoncol.2018.2128] [Medline: 29800065]
4. GBD 2017 Mortality Collaborators. Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1684-1735 [FREE Full text] [doi: 10.1016/S0140-6736(18)31891-9] [Medline: 30496102]
5. Umıt EG, Baysal M, Bas V, Asker I, Kırkızlar O, Demır AM. Polypharmacy and potentially inappropriate medication use in older patients with multiple myeloma, related to fall risk and autonomous neuropathy. J Oncol Pharm Pract 2020 Jan;26(1):43-50. [doi: 10.1177/1078155219835303] [Medline: 30890065]

6. Leger DY, Moreau S, Signol N, Fargeas J, Picat M, Penot A, et al. Polypharmacy, potentially inappropriate medications and drug-drug interactions in geriatric patients with hematologic malignancy: observational single-center study of 122 patients. J Geriatr Oncol 2018 Jan;9(1):60-67. [doi: 10.1016/j.jgo.2017.07.015] [Medline: 28851511]
7. Dhakal P, Lyden E, Muir KE, Al-Kadhimi ZS, Koll T, Maness LJ, et al. Prevalence and effects of polypharmacy on overall survival in acute myeloid leukemia. Leuk Lymphoma 2020 Jul;61(7):1702-1708. [doi: 10.1080/10428194.2020.1737687] [Medline: 32157936]
8. Mian H, Fiala M, Wildes TM. Adherence to lenalidomide in older adults with newly diagnosed multiple myeloma. Clin Lymphoma Myeloma Leuk 2020 Feb;20(2):98-104 [FREE Full text] [doi: 10.1016/j.clml.2019.09.618] [Medline: 31843543]
9. Kristjánsdóttir ER, Toksvang LN, Schmiegelow K, Rank CU. Prevalence of non-adherence and non-compliance during maintenance therapy in adults with acute lymphoblastic leukemia and their associations with survival. Eur J Haematol 2022 Feb;108(2):109-117. [doi: 10.1111/ejh.13711] [Medline: 34562333]
10. Mancini J, Simeoni M, Parola N, Clement A, Vey N, Sirvent N, et al. Adherence to leukemia maintenance therapy: a comparative study among children, adolescents, and adults. Pediatr Hematol Oncol 2012 Aug;29(5):428-439. [doi: 10.3109/08880018.2012.693150] [Medline: 22712832]
11. Hall A, Lynagh M, Bryant J, Sanson-Fisher R. Supportive care needs of hematological cancer survivors: a critical review of the literature. Crit Rev Oncol Hematol 2013 Oct;88(1):102-116. [doi: 10.1016/j.critrevonc.2013.03.008] [Medline: 23582686]
12. Allart-Vorelli P, Porro B, Baguet F, Michel A, Cousson-Gélie F. Haematological cancer and quality of life: a systematic literature review. Blood Cancer J 2015 Apr 24;5:e305 [FREE Full text] [doi: 10.1038/bcj.2015.29] [Medline: 25909835]
13. Allart P, Soubeyran P, Cousson-Gélie F. Are psychosocial factors associated with quality of life in patients with haematological cancer? A critical review of the literature. Psychooncology 2013 Feb;22(2):241-249. [doi: 10.1002/pon.3026] [Medline: 22287503]
14. Phillips JL, Currow DC. Cancer as a chronic disease. Collegian 2010;17(2):47-50. [doi: 10.1016/j.colegn.2010.04.007] [Medline: 20738055]
15. Global Burden of Disease 2019 Cancer Collaboration. Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. JAMA Oncol 2022;8(3):420-444 [FREE Full text] [doi: 10.1001/jamaoncol.2021.6987] [Medline: 34967848]
16. Smartphone market in Europe: statistics and facts. Statista. URL: https://www.statista.com/topics/3341/smartphone-market-in-europe/ [accessed 2022-06-07]
17. Jagannath S, Mikhael J, Nadeem O, Raje N. Digital health for patients with multiple myeloma: an unmet need. JCO Clin Cancer Inform 2021 Oct;5:1096-1105 [FREE Full text] [doi: 10.1200/CCI.20.00145] [Medline: 34735265]
18. Binder AF, Handley NR, Wilde L, Palmisiano N, Lopez AM. Treating hematologic malignancies during a pandemic: utilizing telehealth and digital technology to optimize care. Front Oncol 2020 Jun 26;10:1183 [FREE Full text] [doi: 10.3389/fonc.2020.01183] [Medline: 32676459]
19. Psihogios AM, Li Y, Ahmed A, Huang J, Kersun LS, Schwartz LA, et al. Daily text message assessments of 6-mercaptopurine adherence and its proximal contexts in adolescents and young adults with leukemia: a pilot study. Pediatr Blood Cancer 2021 Feb;68(2):e28767. [doi: 10.1002/pbc.28767] [Medline: 33073479]
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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
22. Dawson RM, Felder TM, Donevant SB, McDonnell KK, Card EB, King CC, et al. What makes a good health 'app'? Identifying the strengths and limitations of existing mobile application evaluation tools. Nurs Inq 2020 Apr;27(2):e12333. [doi: 10.1111/nin.12333] [Medline: 31854055]
23. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
24. Kwan YH, Ong WJ, Xiong M, Leung YY, Phang JK, Wang CTM, et al. Evaluation of mobile apps targeted at patients with spondyloarthritis for disease monitoring: systematic app search. JMIR Mhealth Uhealth 2019 Oct 28;7(10):e14753 [FREE Full text] [doi: 10.2196/14753] [Medline: 31661080]
25. Amor-García, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
26. 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]

27. 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]
28. Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, et al. Quality, functionality, and features of chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e14836 [FREE Full text] [doi: 10.2196/14836] [Medline: 32255432]
29. Escoffery C, McGee R, Bidwell J, Sims C, Thropp EK, Frazier C, et al. A review of mobile apps for epilepsy self-management. Epilepsy Behav 2018 Apr;81:62-69. [doi: 10.1016/j.yebeh.2017.12.010] [Medline: 29494935]
30. Masterson CRM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
31. Collado-Borrell R, Escudero-Vilaplana V, Narrillos-Moraza Á. Patient-reported outcomes y aplicaciones móviles. Revisión de su impacto en los resultados en salud de los pacientes. Farmacia Hospitalaria 2022:9. [doi: 10.7399/fh.11830]
32. Tsatsou I, Konstantinidis T, Kalemikerakis I, Adamakidou T, Vlachou E, Govina O. Unmet supportive care needs of patients with hematological malignancies: a systematic review. Asia Pac J Oncol Nurs 2021;8(1):5-17 [FREE Full text] [doi: 10.4103/apjon.apjon_41_20] [Medline: 33426184]
33. Suleman A, Vijenthira A, Berlin A, Prica A, Rodin D. The use of virtual care in patients with hematologic malignancies: a scoping review. Curr Oncol 2022 Feb 07;29(2):892-900 [FREE Full text] [doi: 10.3390/curroncol29020076] [Medline: 35200575]
34. Taylor S, Bellhouse S, Allsop M, Radford J, Yorke J. The role of eHealth in the delivery of care for patients with hematological cancers: a systematic literature review. Telemed J E Health 2020 Sep;26(9):1093-1105. [doi: 10.1089/tmj.2019.0231] [Medline: 32208067]
35. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
36. Adam R, McMichael D, Powell D, Murchie P. Publicly available apps for cancer survivors: a scoping review. BMJ Open 2019 Sep 30;9(9):e032510 [FREE Full text] [doi: 10.1136/bmjopen-2019-032510] [Medline: 31575584]
37. 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. [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
38. Hummelinck A, Pollock K. Parents' information needs about the treatment of their chronically ill child: a qualitative study. Patient Educ Couns 2006 Aug;62(2):228-234. [doi: 10.1016/j.pec.2005.07.006] [Medline: 16139981]
39. Modi AC, Pai AL, Hommel KA, Hood KK, Cortina S, Hilliard ME, et al. Pediatric self-management: a framework for research, practice, and policy. Pediatrics 2012 Feb;129(2):e473-e485 [FREE Full text] [doi: 10.1542/peds.2011-1635] [Medline: 22218838]
40. Owens OL, Beer JM, Reyes LI, Thomas TL. Systematic review of commercially available mobile phone applications for prostate cancer education. Am J Mens Health 2019;13(1):1557988318816912 [FREE Full text] [doi: 10.1177/1557988318816912] [Medline: 30526243]
41. Stolwijk C, van Onna M, Boonen A, van Tubergen A. Global prevalence of spondyloarthritis: a systematic review and meta-regression analysis. Arthritis Care Res (Hoboken) 2016 Sep;68(9):1320-1331 [FREE Full text] [doi: 10.1002/acr.22831] [Medline: 26713432]
42. Cannon C. Telehealth, mobile applications, and wearable devices are expanding cancer care beyond walls. Semin Oncol Nurs 2018 May;34(2):118-125. [doi: 10.1016/j.soncn.2018.03.002] [Medline: 29627143]
43. Smith A, Howell D, Patmore R, Jack A, Roman E. Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. Br J Cancer 2011 Nov 22;105(11):1684-1692 [FREE Full text] [doi: 10.1038/bjc.2011.450] [Medline: 22045184]
44. Tietsche de Moraes Hungria V, Chiattone C, Pavlovsky M, Abenoza LM, Agreda GP, Armenta J, et al. Epidemiology of hematologic malignancies in real-world settings: findings from the hemato-oncology latin america observational registry study. J Glob Oncol 2019 Nov;5:1-19 [FREE Full text] [doi: 10.1200/JGO.19.00025] [Medline: 31774711]
45. Heponiemi T, Kaihlanen A, Kouvonen A, Leemann L, Taipale S, Gluschkoff K. The role of age and digital competence on the use of online health and social care services: a cross-sectional population-based survey. Digit Health 2022;8:20552076221074485 [FREE Full text] [doi: 10.1177/20552076221074485] [Medline: 35111333]
46. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
47. Nieman CL, Oh ES. Connecting with older adults via telemedicine. Ann Intern Med 2020 Nov 17;173(10):831-832. [doi: 10.7326/M20-1322] [Medline: 32777187]
48. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]

49. Medicine evaluation figures. European Medicines Agency. URL: https://www.ema.europa.eu/en/about-us/what-we-do/authorisation-medicines/medicine-evaluation-figures [accessed 2022-06-09]
50. Finnes HD. 2019-2020 drug updates in hematologic malignancies. J Adv Pract Oncol 2021 Apr;12(3):279-283 [FREE Full text] [doi: 10.6004/jadpro.2021.12.3.12] [Medline: 34084575]
51. Wu Y, Linder L, Kanokvimankul P, Fowler B, Parsons B, Macpherson C, et al. Use of a smartphone application for prompting oral medication adherence among adolescents and young adults with cancer. Oncol Nurs Forum 2018 Jan 01;45(1):69-76 [FREE Full text] [doi: 10.1188/18.ONF.69-76] [Medline: 29251285]
52. Heneghan MB, Hussain T, Barrera L, Cai SW, Haugen M, Morgan E, et al. Access to technology and preferences for an mhealth intervention to promote medication adherence in pediatric acute lymphoblastic leukemia: approach leveraging behavior change techniques. J Med Internet Res 2021 Feb 18;23(2):e24893 [FREE Full text] [doi: 10.2196/24893] [Medline: 33599621]
53. Pereira-Salgado A, Westwood JA, Russell L, Ugalde A, Ortlepp B, Seymour JF, et al. Mobile health intervention to increase oral cancer therapy adherence in patients with chronic myeloid leukemia (The REMIND System): clinical feasibility and acceptability assessment. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e184 [FREE Full text] [doi: 10.2196/mhealth.8349] [Medline: 29212628]
54. Raghunathan NJ, Korenstein D, Li QS, Tonorezos ES, Mao JJ. Determinants of mobile technology use and smartphone application interest in cancer patients. Cancer Med 2018 Nov;7(11):5812-5819 [FREE Full text] [doi: 10.1002/cam4.1660] [Medline: 30280495]
55. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
56. Tarricone R, Cucciniello M, Armeni P, Petracca F, Desouza KC, Hall LK, et al. Mobile health divide between clinicians and patients in cancer care: results from a cross-sectional international survey. JMIR Mhealth Uhealth 2019 Sep 06;7(9):e13584 [FREE Full text] [doi: 10.2196/13584] [Medline: 31493318]
57. Nair SC, Hassen WA, Sreedharan J, Qawasmeh K, Ibrahim H. Comparison of perceptions of unmet supportive care needs between cancer patients and their oncologists. Support Care Cancer 2019 Nov;27(11):4229-4235. [doi: 10.1007/s00520-019-04705-4] [Medline: 30847701]
58. Subhi Y, Bube SH, Rolskov BS, Skou TAS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
59. Haffey F, Brady RRW, Maxwell S. A comparison of the reliability of smartphone apps for opioid conversion. Drug Saf 2013 Feb;36(2):111-117. [doi: 10.1007/s40264-013-0015-0] [Medline: 23322549]
60. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
61. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
62. Connor K, Brady RRW, de Beaux A, Tulloh B. Contemporary hernia smartphone applications (apps). Hernia 2014 Aug;18(4):557-561. [doi: 10.1007/s10029-013-1130-7] [Medline: 23801277]
63. Mair FS, Montori VM, May CR. Digital transformation could increase the burden of treatment on patients. BMJ 2021 Nov 25;375:n2909. [doi: 10.1136/bmj.n2909] [Medline: 34824093]
64. Mair FS, May CR. Thinking about the burden of treatment. BMJ 2014 Nov 10;349:g6680. [doi: 10.1136/bmj.g6680] [Medline: 25385748]
65. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
66. Biggest app stores in the world. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2022-06-11]

Abbreviations

HM: hematological malignancy
ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for a Systematic Review and Meta-analysis

Edited by L Buis; submitted 20.12.21; peer-reviewed by R Halkes, L Guo; comments to author 06.05.22; revised version received 21.06.22; accepted 02.09.22; published 20.09.22.

Please cite as:
Báez Gutiérrez N, Rodríguez Ramallo H, Fernández González M, Abdel-Kader Martín L
Smartphone Apps for Patients With Hematologic Malignancies: Systematic Review and Evaluation of Content
JMIR Mhealth Uhealth 2022;10(9):e35851
URL: https://mhealth.jmir.org/2022/9/e35851
doi:10.2196/35851
PMID:36125860

Original Paper

Mobile Health Use by Older Individuals at Risk of Cardiovascular Disease and Type 2 Diabetes Mellitus in an Australian Cohort: Cross-sectional Survey Study

Vera Helen Buss[1,2], PharmB, RPh, MRes; Marlien Varnfield[1], BSc, MSc, PhD; Mark Harris[2], MBBS, MD; Margo Barr[2], BSc, MPH, PhD

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Herston, Australia

[2]Centre for Primary Health Care and Equity, University of New South Wales, Sydney, Australia

Corresponding Author:
Vera Helen Buss, PharmB, RPh, MRes
Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
Level 7, Surgical Treatment and Rehabilitation Service
296 Herston Road
Herston, 4029
Australia
Phone: 61 7 3253 3603
Email: vera.buss@csiro.au

Abstract

Background: The digital transformation has the potential to change health care toward more consumers' involvement, for example, in the form of health-related apps which are already widely available through app stores. These could be useful in helping people understand their risk of chronic conditions and helping them to live more healthily.

Objective: With this study, we assessed mobile health app use among older Australians in general and among those who were at risk of cardiovascular disease or type 2 diabetes mellitus.

Methods: In this cross-sectional analysis, we used data from the second follow-up wave of the 45 and Up Study. It is a cohort study from New South Wales, Australia, with 267,153 participants aged 45 years and older that is based on a random sample from the Services Australia (formerly the Australian Government Department of Human Services) Medicare enrollment database. The 2019 follow-up questionnaire contained questions about technology and mobile health use. We further used data on prescribed drugs and hospitalizations to identify participants who already had cardiovascular disease or diabetes or who were at risk of these conditions. Our primary outcome measure was mobile health use, defined as having used a mobile health app before. We used descriptive statistics and multivariate logistic regression to answer the research questions.

Results: Overall, 31,946 individuals with a median age of 69 (IQR 63-76) years had completed the follow-up questionnaire in 2019. We classified half (16,422/31,946, 51.41%) of these as being at risk of cardiovascular disease or type 2 diabetes mellitus and 38.04% (12,152/31,946) as having cardiovascular disease or type 1 or type 2 diabetes mellitus. The proportion of mobile health app users among the at-risk group was 31.46% (5166/16,422) compared to 29.16% (9314/31,946) in the total sample. Those who used mobile health apps were more likely to be female, younger, without physical disability, and with a higher income. People at risk of cardiovascular disease or type 2 diabetes mellitus were not statistically significantly more likely to use mobile health than were people without risk (odds ratio 1.06, 95% CI 0.97-1.16; *P*=.18; adjusted for age, sex, income, and physical disability).

Conclusions: People at risk of cardiovascular disease or type 2 diabetes mellitus were not more likely to use mobile health apps than were people without risk. Those who used mobile health apps were less likely to be male, older, with a physical disability, and with a lower income. From the results, we concluded that aspects of equity must be considered when implementing a mobile health intervention to reach all those that can potentially benefit from it.

(JMIR Mhealth Uhealth 2022;10(9):e37343) doi:10.2196/37343

KEYWORDS

cardiovascular diseases; diabetes mellitus type 2; cohort studies; telemedicine; mobile applications; mHealth; mobile health; mobile app; aging; digital health; cardiovascular; diabetes

Introduction

Chronic conditions pose a great burden to health systems [1]. One of the main drivers for the high prevalence of chronic diseases such as cardiovascular disease (CVD) and type 2 diabetes mellitus (T2DM) is an unhealthy lifestyle [2]. This includes smoking, high alcohol consumption, a lack of physical activity, and a poor diet [2]. Health promotion campaigns aim at increased awareness for lifestyle-related risk factors and support for risk reduction [3]. Digital applications, such as mobile health (mHealth) apps for smartphones, offer an excellent opportunity to make such health promotion interventions available to as many people as possible [3]. A report by the IQVIA Institute stated that in 2021, over 350,000 health-related apps were available through app stores [4]. According to the report, 90,000 new health-related apps appeared on the market in 2020 alone [4]. The mHealth space is not limited to apps; other products include smart watches and activity trackers, commonly called wearables [5].

Meskó et al [6] called digital health the cultural transformation of traditional health care. In a recent position paper by the European Society of Cardiology Working Group on e-cardiology [5], the authors mentioned that the advantages of wearables include the democratization of health data, the potential for earlier detection of risk factors and disease states, and increased health awareness. The authors also noted that while older and high-risk individuals would likely benefit the most, mHealth technology is currently used the most by the younger, healthy generation with higher income [5]. More generally, Spoth et al [7] reported that those who are most in need of evidence-based prevention and health promotion interventions often do not use them, in many cases due to lack of access. To understand the possible uptake of an app-based preventive intervention for CVD and T2DM that we are currently developing, we examined mHealth use among older Australians. Previous research has reported on mHealth uptake in other countries. For example, Shan et al [8], Robbins et al [9], Rising et al [10], and Carroll et al [11] conducted studies using US data; Pare et al [12] analyzed Canadian data; and Xi et al [13] investigated mHealth uptake in Chinese adults. However, so far there has not been a larger study looking at mHealth use in the Australian setting. For this purpose, we analyzed data on mHealth use collected in 2019 from the second follow-up wave of the 45 and Up Study which is a large cohort study of Australians who were aged 45 years and older at baseline between 2006 and 2009 [14,15]. With this study, we also aimed to identify any distinguishing features between those who used mHealth and those who did not. Our specific research questions were as follows: What was the proportion of mHealth users overall and among those who were at risk of CVD or T2DM? How did mHealth users differ from non-mHealth users for people at risk of CVD or T2DM? Were those at risk of CVD or T2DM more likely to use mHealth than those without risk?

Methods

Ethics Approval

This cross-sectional study was based on survey data from the Sax Institute's 45 and Up Study. This is a large cohort including 267,153 participants aged 45 years and older who reside in New South Wales (NSW), Australia [16]. The conduct of the 45 and Up Study was approved by the University of New South Wales Human Research Ethics Committee. The survey data were linked to data from the Pharmaceutical Benefits Scheme (PBS; data on prescribed drugs) and the NSW Admitted Patient Data Collection (APDC; hospitalization data). This study has been approved by the NSW Population & Health Services Research Ethics Committee (approval #HREC/16/CIPHS/14) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Health and Medical Human Research Ethics Committee (approval #2021_018_RR). All cohort participants provided free and informed consent.

Data Sources

The recruitment for the 45 and Up Study took place between January 2006 and December 2009 and was based on a random sample from the Services Australia (formerly the Australian Government Department of Human Services) Medicare enrollment database, with oversampling in people aged 80 years and over and residents of rural and remote areas [16]. About 18% of those who were contacted consented to take part in the study which represents 11% of the NSW population aged 45 years and older [14]. A detailed description of the cohort and the study methods was published by the 45 and Up Study collaborators [14,15]. For our analyses, we used data from the second wave of follow-up collected between 2018 and 2020, specifically, the 2019 follow-up survey, which included questions about participants' technology and mHealth use [17]. The survey was sent to 68,349 participants, of whom 31,965 responded (46.77%) [16]. In general, the questionnaires contained questions on lifestyle, medical history, family history of chronic conditions, socioeconomic status, and geographic factors [17]. The Sax Institute linked the survey data to the PBS data deterministically using a unique identifier [16]. Additionally, the Centre for Health Record Linkage linked these data to the APDC data using probabilistic techniques [18].

Subgroups

We first identified participants with CVD or diabetes (type 1 or 2, not including gestational diabetes) at the time of the follow-up survey and then identified those who were at risk of the conditions. These methods were similar to the methods described by Joshy et al [19] for the classification of CVD and by Comino et al [20] for the classification of diabetes in the 45 and Up Study. Individuals were classified based on information from the survey, from the APDC data, and from the PBS data. In the survey, participants were asked if they had CVD or diabetes ("Has a doctor ever told you that you have:" heart failure, atrial fibrillation, other heart disease, stroke, OR

diabetes—type 1, type 2 or unsure) and if they took medication for these conditions (corresponding medication listed on the questionnaire: Cardizem/Vasocardol, warfarin, OR Diabex/Diaformin). From the APDC data, the diagnoses were classified based on related hospital admissions before the date of the survey. We searched for relevant International Classification of Diseases version 10 Australian Modification (ICD-10-AM) diagnosis codes in any of the 55 diagnostic fields or relevant Australian Classification of Health Interventions procedure codes in any of the 50 procedure code fields. These codes were based on methods described by the Australian Institute of Health and Welfare (AIHW) [21-24]. From the PBS data, people were classified with CVD or diabetes if they took medication with these indications in the past 12 months [25,26]. For CVD, the only drugs considered were those that are solely indicated for CVD. To identify a CVD diagnosis, we did not consider drugs for hypertension or dyslipidemia. People with hypertension or dyslipidemia may be at risk of CVD but have not yet developed the condition. Additionally, people may take antihypertensive drugs for other reasons than hypertension. Therefore, if people took antihypertensive or lipid-modifying drugs but did not have a diagnosis of CVD (ie, reported in the 45 and Up Study data, the APDC data, or due to taking a drug with CVD as the only indication), we did not classify them as having the condition.

For CVD, being at risk was classified as taking antihypertensive medication, lipid-modifying medication, or low-dose aspirin; reporting blood-clotting problems or hypertension; being obese (≥30 kg/m^2); or having a family history of CVD (parents or siblings) and not already having CVD. For T2DM, someone was classified to be at risk if they had had gestational diabetes, were obese (≥30 kg/m^2), or had a family history of T2DM (parents or siblings) but were free of type 1 or T2DM. We gathered the information to identify at-risk populations from the survey and the PBS data. Women were classified as having had gestational diabetes if they self-reported having had the condition or if they had received a diabetes diagnosis before the date of their last delivery and if there was no evidence of diabetes medication in their records for the 12 months before filling out the follow-up survey.

Technology and mHealth Use

The primary outcome was mHealth use. We defined mHealth use through the following survey question: “How often do you use apps on your mobile phone or tablet to track the following?” The answer options were “never,” “less than once a month,” “at least once a month,” “at least once a week,” or “every day.” If any of the options were selected with “less than once a month” or more frequently, the person was classified as a mHealth user (options included the following: activity or fitness, vital signs, nutrition or weight, mood or well-being, sleep, medications). This meant that if “never” or none of the options was selected, the participant was considered a nonuser. Table 1 summarizes the technology-related questions from the 45 and Up Study questionnaire that we used for the analysis [17]. We classified all those who did not select yes for the questions about device use (computer or laptop, tablet, smartphone, fitness tracker, and smart watch) as not using them. We categorized everyone as app users who did not select “I don’t use apps/don’t know what apps are.” We dichotomized the variable to app download by putting app nonusers in one group with those who selected that they had not downloaded any apps yet and the remaining in the other group.

Table 1. Technology-related questions from the 45 and Up Study questionnaire 2019 that were used in the analysis.

Question	Answer options
Do you use a computer or mobile device (eg, phone with a touch screen, tablet, or smart watch) regularly?	Yes/no
If YES, which of the following devices do you use regularly? Apple desktop or laptop computer (eg, iMac, MacBook)/Windows desktop or laptop computer/Apple tablet (iPad)/other tablet (eg, Samsung, Microsoft Surface, Lenovo)/Apple phone with a touch screen (iPhone)/Android phone with a touch screen (eg, Samsung, Huawei, Google)/other phone with a touch screen/Apple Watch/other smart watch/fitness tracker (eg, Fitbit, Garmin)	Yes/no/unsure
If you use applications (apps) on a mobile phone or tablet, how many have you ever downloaded yourself? (choose one only)	I don’t use apps/don’t know what apps are/none/1-5 /6 or more
How often do you use apps on your mobile phone or tablet to track the following: activity or fitness (eg, number of steps, exercise)/vital signs (eg, heart rate, blood pressure, breathing)/nutrition or weight/mood or well-being/sleep/medications (eg, reminders, alerts)	Never/less than once a month/at least once a month/at least once a week/every day

Other Measures

Other variables of interest included age, sex, income, lifestyle (smoking, alcohol, fruit or vegetables, physical activity), and physical disability. We categorized income based on the Organization for Economic Cooperation and Development (OECD) report [27] into low (less than Aus $30,000 per year [US $20,580]), middle (Aus $30,000-89,999 per year [US $20,580-61,740]), and high income (Aus $90,000 or more per year [US $61,741 or more]). For smoking, we created three categories: never, past, and current smokers. We classified physical disability through survey questions about illness or disability restricting physical activity (“Is there anything that stops you from participating in physical activity?”—ill-health; “Do you regularly need help with daily tasks because of long-term illness or disability?”—yes; “Does your health now limit you in any of the following activities?”—yes selected for any of the options). We dichotomized all other lifestyle-related variables depending on if participants met the guideline recommendations (for alcohol: maximum 10 standard drinks per week [28]; for fruits and vegetables: at least 2 servings of fruits and 5 servings of vegetables per day [29]; for physical activity: at least 150 minutes of physical activity per week with vigorous physical activity counting double [30]). For fruits and

vegetables, we set servings per day to 0 if the participants selected that they did not eat any fruits or vegetables, respectively. Otherwise, we did not impute any missing values. We reported on the percentage of missingness for age, sex, BMI, smoking status, alcohol consumption, fruit and vegetable intake, physical activity, income, and physical disability. For the remaining variables, we did not report on missing values because the way we categorized these variables did not result in any missing values.

Statistical Analysis

We used descriptive statistics to describe the demographics of the overall population and the various subgroups (with CVD or diabetes; at risk of CVD or T2DM; free of CVD, T2DM, and their risk factors) as well as their technology and mHealth use. Further, we used the *t* test and the chi-square test to check for differences between mHealth users and nonusers among those at risk of CVD or T2DM. We defined statistical significance at a level of 5%. We built multivariable logistic regression models for the total sample and those at risk of CVD or T2DM to assess the influence of age, sex, physical disability, and income (predictor variables) on mHealth use (outcome variable). From these models, we calculated adjusted odds ratios (ORs) and their 95% CIs. Further, we computed the adjusted OR of mHealth use for someone with CVD or T2DM risk compared to someone free of the condition and not at risk. We conducted the analyses in RStudio (version 1.2.5042) using the programming language R (version 4.0.0; R Foundation for Statistical Computing) within the Secured Unified Research Environment (SURE) provided by the Sax Institute [31].

Results

In total, 31,946 individuals completed the follow-up questionnaire in 2019 that contained the mHealth questions. From these 31,946 participants, 12,152 (38.04%) participants had CVD or diabetes at the time of the follow-up survey, 16,422 (51.41%) participants were categorized as at risk of the conditions, and 3372 (10.56%) participants were categorized as not at risk (Figure 1). The variable with the highest proportion of missing values was fruit and vegetable intake with 7.66% (2447/31,946; Table 2).

Table 3 compares the characteristics of the total sample and the various subgroups, including information on device ownership and app use. The group with CVD or diabetes included older people, more men, more people with physical disabilities, more people who were obese, and more exsmokers than did the at-risk group. The difference for these variables was even greater when compared to the group without CVD, diabetes, and risk factors. Device ownership was higher in people without CVD and diabetes. Overall, 75.93% (24,256/31,946) of the total sample stated that they owned a smartphone and 29.16% (9314/31,946) of the total sample stated that they had used a health app before. The highest proportion was in the group without conditions and risk factors (1102/3372, 32.68%), followed by the at-risk group (5166/16,422, 31.46%), and last the group with CVD or diabetes (3046/12,152, 25.07%).

Figure 1. Flowchart of study participants. CVD: cardiovascular disease; mHealth: mobile health.

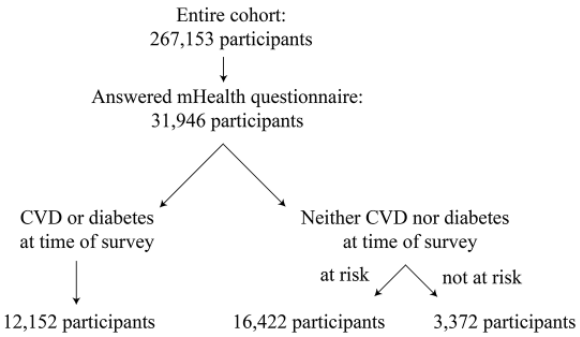


Table 2. Proportion of missing values for each variable of interest (N=31,946).

Variable	Missing, n (%)
Age	0 (0)
Sex	0 (0)
BMI	1528 (4.78)
Smoking status	282 (0.88)
Alcohol consumption	697 (2.18)
Fruit and vegetable intake	2447 (7.66)
Physical activity	964 (2.96)
Income	1303 (4.08)
Physical disability	1745 (5.46)

Table 3. Demographic characteristics of all participants and various subgroups.

Characteristics	Total sample	With CVD[a] or diabetes	At risk of CVD or T2DM[b]	Free of CVD and diabetes and not at risk
Subgroup, n (% of total sample)	31,946 (100)	12,152 (38.04)	16,422 (51.41)	3372 (10.56)
Age, median (IQR)	69 (63-76)	73 (67-79)	67 (62-3)	65 (61-71)
Female, n (%)	16,462 (51.53)	5181 (42.63)	9517 (57.95)	1764 (52.31)
Physical disability, n (%)	14,963 (49.54)	7598 (65.60)	6433 (41.56)	932 (29.68)
Weight, n (%)				
Normal	10,677 (35.10)	3302 (28.69)	5558 (35.39)	1817 (56.69)
Overweight	11,851 (38.96)	4475 (38.88)	5988 (38.13)	1388 (43.31)
Obese	7890 (25.94)	3733 (32.43)	4157 (26.47)	0 (0)
Smoking status, n (%)				
Never	19,259 (60.82)	6801 (56.55)	10,288 (63.15)	2170 (64.85)
Past	11,508 (36.34)	4924 (40.94)	5528 (33.93)	1056 (31.56)
Current	897 (2.83)	302 (2.51)	475 (2.92)	120 (3.59)
Meeting guideline recommendations, n (%)				
Alcohol	24,971 (79.9)	9621 (81.62)	12,760 (79.02)	2590 (78.13)
Fruits and vegetables	7039 (23.9)	2572 (23.26)	3744 (24.40)	723 (23.35)
Physical activity	24,025 (77.5)	8111 (69.26)	13,159 (81.97)	2755 (85.61)
Income, n (%)				
Low	6377 (20.8)	3128 (27.16)	2770 (17.44)	479 (14.74)
Middle	12,789 (41.7)	4705 (40.86)	6732 (42.40)	1352 (41.61)
High	6486 (21.2)	1810 (15.72)	3769 (23.74)	907 (27.92)
Prefer not to say	4991 (16.3)	1872 (16.26)	2608 (16.42)	511 (15.73)
Device ownership, n (%)				
Laptop or computer	22,610 (70.78)	7891 (64.94)	12,160 (74.05)	2559 (75.89)
Tablet	15,368 (48.11)	5209 (42.87)	8506 (51.80)	1653 (49.02)
Smartphone	24,256 (75.93)	8198 (67.46)	13,319 (81.10)	2739 (81.23)
Fitness tracker	3523 (11.03)	1118 (9.20)	1999 (12.17)	406 (12.04)
Smart watch	1506 (4.71)	513 (4.22)	811 (4.94)	182 (5.40)
App use, n (%)				
Any app	26,434 (82.75)	9468 (77.91)	14,050 (85.56)	2916 (86.48)
Downloading apps	22,336 (69.92)	7763 (63.88)	12,042 (73.33)	2540 (75.33)
Health apps	9314 (29.16)	3046 (25.07)	5166 (31.46)	1102 (32.68)

[a]CVD: cardiovascular disease.

[b]T2DM: type 2 diabetes mellitus.

Of those who used mHealth, physical activity was the most tracked feature (Table 4). Overall, 24.73% of participants (7900/31,946) had tracked their physical activity levels with an app before. Among those who were at risk of CVD or T2DM, the proportion was slightly higher at 27.20% (4467/16,422). The second most-tracked feature was vital signs (total: 7315/31,946, 22.90%; at risk: 4168/16,422, 25.38%). All other features were tracked by less than 10% of the sample. Among those who stated that they tracked physical activity, most did so daily (total: 3600/31,946, 11.27%; at risk: 2032/16,422, 12.37%). The same tendency was observed for tracking medication, sleep, and vital signs. For mood and weight or diet, those who stated that they tracked the feature were mostly reporting a frequency of less than once a month.

We observed statistically significant differences between mHealth users and nonusers among the at-risk population for the variables sex, hypertension, gestational diabetes, family history, physical disability, weight, smoking status, meeting physical activity guideline recommendations, and owning technical devices (Table 5).

When adjusted for other demographic factors, younger participants, women, people without physical disability, and people with higher income were more likely to use mHealth (Figure 2 for all participants and Figure 3 for the at-risk population). The adjusted OR of using mHealth for someone who was at risk of CVD or T2DM compared to someone who was not at risk and free of both conditions was 1.06 (95% CI 0.97-1.16; *P*=.18; adjusted for age, sex, income, and physical disability).

Table 4. Frequency of mHealth use (apps with health-related tracking features) overall and among the subgroups.

Sample by frequency of tracking	Physical activity, n (%)[a]	Medication, n (%)[a]	Mood, n (%)[a]	Weight or diet, n (%)[a]	Sleep, n (%)[a]	Vital signs, n (%)[a]
Total						
Never	24,046 (75.27)	30,552 (95.64)	31,133 (97.46)	29,344 (91.86)	29,418 (92.09)	24,631 (77.10)
<1/month	1314 (4.11)	302 (0.95)	389 (1.22)	828 (2.59)	485 (1.52)	1254 (3.93)
Monthly	790 (2.47)	217 (0.68)	130 (0.41)	460 (1.44)	307 (0.96)	731 (2.29)
Weekly	2,196 (6.87)	247 (0.77)	156 (0.49)	716 (2.24)	572 (1.79)	1983 (6.21)
Daily	3,600 (11.27)	628 (1.97)	138 (0.43)	598 (1.87)	1,164 (3.64)	3347 (10.48)
With CVD[b] or diabetes						
Never	9696 (79.80)	11,480 (94.47)	11,906 (97.98)	11,297 (92.96)	11,361 (93.49)	9921 (81.64)
<1/month	404 (3.32)	124 (1.02)	105 (0.86)	243 (2.00)	132 (1.09)	377 (3.10)
Monthly	240 (1.97)	96 (0.79)	45 (0.37)	167 (1.37)	91 (0.75)	217 (1.79)
Weekly	679 (5.59)	117 (0.96)	47 (0.39)	228 (1.87)	183 (1.51)	592 (4.87)
Daily	1,133 (9.32)	335 (2.76)	49 (0.40)	217 (1.79)	385 (3.17)	1045 (8.60)
At risk of CVD or T2DM[c]						
Never	11,955 (72.80)	15,806 (96.25)	15,949 (97.12)	14,940 (90.98)	14,966 (91.13)	12,254 (74.62)
<1/month	739 (4.50)	152 (0.93)	236 (1.44)	495 (3.01)	294 (1.79)	712 (4.34)
Monthly	459 (2.80)	103 (0.63)	68 (0.41)	246 (1.50)	182 (1.11)	428 (2.61)
Weekly	1237 (7.53)	106 (0.64)	90 (0.55)	411 (2.50)	313 (1.91)	1135 (6.91)
Daily	2032 (12.37)	255 (1.55)	79 (0.48)	330 (2.01)	667 (4.06)	1893 (11.53)
Free of CVD and diabetes and not at risk						
Never	2395 (71.03)	3266 (96.86)	3278 (97.21)	3107 (92.14)	3091 (91.67)	2456 (72.84)
<1/month	171 (5.07)	26 (0.77)	48 (1.42)	90 (2.67)	59 (1.75)	165 (4.89)
Monthly	91 (2.70)	18 (0.53)	17 (0.50)	47 (1.39)	34 (1.01)	86 (2.55)
Weekly	280 (8.30)	24 (0.71)	19 (0.56)	77 (2.28)	76 (2.25)	256 (7.59)
Daily	435 (12.90)	38 (1.13)	10 (0.30)	51 (1.51)	112 (3.32)	409 (12.13)

[a]Percentages may not total 100% due to rounding.

[b]CVD: cardiovascular disease.

[c]T2DM: type 2 diabetes mellitus.

Table 5. Differences between participants at risk of CVD and/or T2DM who use mHealth and who do not.

Characteristics	At risk of CVD[a] or T2DM[b]			
	mHealth[c] users, n (%)[d]	Nonusers, n (%)[d]	Chi-square (*df*)[e]	*P* value[e]
Sample size	5166 (31.46)	11,256 (68.54)	N/A[f]	N/A
Age, median (IQR) [g]	64 (60-69)	69 (63-75)	N/A	.47
Female	3177 (61.50)	6340 (56.33)	38.7 (1)	<.001
Hypertension	2199 (42.57)	5355 (47.57)	35.5 (1)	<.001
Dyslipidemia	1539 (29.79)	3462 (30.76)	1.5 (1)	.22
Gestational diabetes, n (% of women)	46 (1.45)	56 (0.88)	5.8 (1)	.02
Family history of CVD	3564 (68.99)	7488 (66.52)	9.7 (1)	.002
Family history of diabetes	1421 (27.51)	2653 (23.57)	29.2 (1)	<.001
Physical disability	1611 (32.80)	4822 (45.63)	226.5 (1)	<.001
Weight			9.8 (2)	.01
Normal	1690 (33.73)	3868 (36.18)		
Overweight	1981 (39.53)	4007 (37.48)		
Obese	1340 (26.74)	2817 (26.35)		
Smoking status			38.4 (2)	<.001
Never	3225 (62.71)	7063 (63.36)		
Past	1826 (35.50)	3702 (33.21)		
Current	92 (1.79)	383 (3.44)		
Following guideline recommendations				
Alcohol	4058 (79.29)	8702 (78.90)	0.3 (1)	.59
Fruits and vegetables	1225 (24.85)	2519 (24.18)	0.8 (1)	.38
Physical activity	4505 (88.59)	8654 (78.90)	220.2 (1)	<.001
Income			626.4 (3)	<.001
Low	500 (9.89)	2270 (20.98)		
Middle	2,034 (40.22)	4698 (43.41)		
High	1750 (34.61)	2019 (18.66)		
Prefer not to say	773 (15.29)	1835 (16.96)		
Device ownership				
Laptop or computer	4394 (85.06)	7766 (68.99)	474.5 (1)	<.001
Tablet	3510 (67.94)	4996 (44.39)	786.2 (1)	<.001
Smartphone	4982 (96.44)	8337 (74.07)	1154.9 (1)	<.001
Fitness tracker	1694 (32.79)	305 (2.71)	2994.3 (1)	<.001
Smart watch	686 (13.28)	125 (1.11)	1114.3 (1)	<.001

[a]CVD: cardiovascular disease.

[b]T2DM: type 2 diabetes mellitus.

[c]mHealth: mobile health.

[d]Age is presented as median (IQR).

[e]For age: Wilcoxon rank-sum test with continuity correction; for all other variables: Pearson chi-square test with Yates continuity correction.

[f]N/A: not applicable.

[g]W=3.

Figure 2. Forest plot with adjusted odds ratios for using mHealth in the entire cohort. mHealth: mobile health.

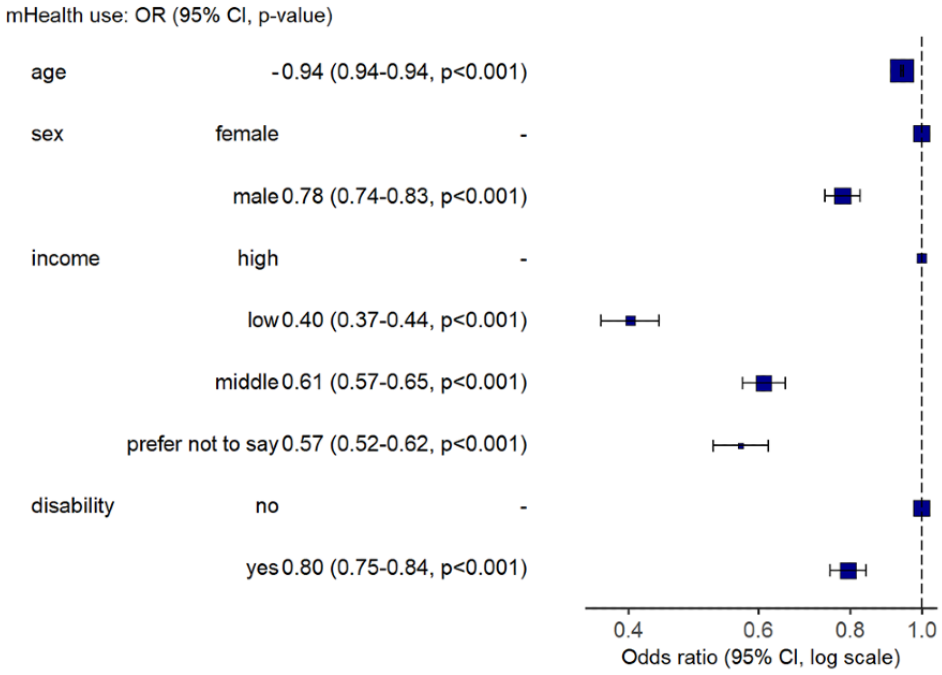
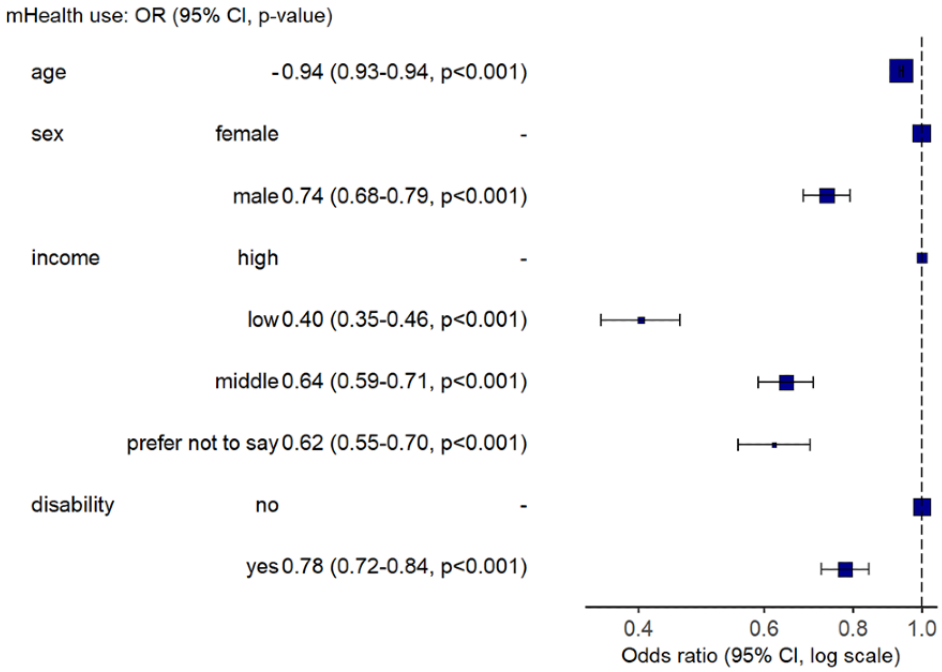


Figure 3. Forest plot with adjusted odds ratios for using mHealth in those at risk of cardiovascular disease or type 2 diabetes mellitus. mHealth: mobile health.



Discussion

Principal Findings

The overall aim of this analysis was to understand how older Australians in general and particularly those at risk of CVD or T2DM use mHealth. To our knowledge, this is the first study of this kind in Australia. Among the at-risk population, the proportion of mHealth users was slightly higher than that of the general proportion. The multivariable logistic regression analysis showed that women, younger people, individuals without disabilities, and higher earners had higher odds of using mHealth. Among the mHealth users, there were fewer smokers and fewer people with hypertension or physical disability. On the other hand, among those who did not use mHealth, fewer people were overweight and fewer reported a family history of CVD or T2DM. According to our results, people at risk of CVD or T2DM do not have higher odds of using mHealth than do those without risk.

Comparison to Other Work

The results largely corresponded to the results of other researchers. Shan et al [8] compared mHealth use among US Americans with CVD or at risk with those without based on

data from 2018. They concluded that those who had or were at risk of CVD were less likely to use mHealth than were the rest of the study population in an unadjusted comparison. However, when adjusted for age, race, education, household income, health insurance, and urban or rural location, the OR of having a health app on the smartphone was 1.24 (95% CI 0.85-1.81; *P*=.26) for women with CVD or at risk of CVD and 1.12 (95% CI 0.68-1.84; *P*=.65) for men with CVD or at risk of CVD compared to women and men with no history or risk factors of CVD. Seifert and Vandelanotte [32] reported on the use of wearables and mHealth apps among Swiss adults aged 65 years and older based on data from 2019. Of their 1149 participants, 43.1% owned a tablet, 68.7% owned a smartphone, 7.6% owned a fitness tracker, 3.3% owned a smart watch, and 22.9% used mHealth apps. Bhuyan et al [33] analyzed 2014 data from US American adults on mHealth use and health-oriented behavior. They found that 35.9% of the smartphone or tablet owners had mHealth apps installed. The proportion of those who had mHealth apps was much lower among those aged 55 to 64 years (10.6%) and those aged 65 years and over (4.7%). Robbins et al [9] analyzed 2015 survey data on mHealth use among US American adults who owned a mobile phone. Among the participants, daily mHealth use was more common in individuals without diseases (21.3%) than in people with high blood pressure (2.7%), with obesity (13.1%), with diabetes (12.3%), with depression (12.0%), or with high cholesterol (16.6%). Robbins et al [9] concluded that their findings suggest that those most likely to benefit from mHealth are least likely to use it.

Limitations

Only 18% of those who were invited to participate in the 45 and Up Study took the baseline survey [14]. This is comparable to other large international cohort studies [34-36]. Mealing et al [37] conducted an analysis in which they compared exposure-outcome relationships in the 45 and Up Study to the NSW Population Health Survey which had a response rate of about 60% [38]. Both are based on the same population, and the analysis showed that the results from both cohorts can be generalized. Additionally, some participants were lost to follow-up, which might also raise concerns about generalizability. Wang et al [39] built a logistic regression model to assess the influence of nonresponse in the first follow-up survey of the 45 and Up Study. In their conclusion, they reported that it did not lead to substantial bias and essentially did not change the interpretation of the results [39]. Further limitations were that the data set had missing values, especially concerning fruit and vegetable intake, and was based on self-report. Ng et al [40] considered the bias through self-reported weight and height in the 45 and Up Study. They stated that the resulting BMI values were logical but underestimated being overweight or obese [40]. There is also a risk of misclassification of disease diagnosis or risk. To minimize the risk, we applied methods that have been described by the AIHW [22-24] and by other researchers who worked with the 45 and Up Study [19,20]. In our analysis, we did not classify diabetes as a risk factor for CVD because we categorized people with CVD or T2DM as a separate subgroup. There is a small risk of wrongly classifying people as at risk of CVD because they took antihypertensive drugs for an indication unrelated to an increased CVD risk.

Implications

The availability of mHealth keeps increasing, and there is little doubt that it can positively impact health promotion, risk awareness, early detection, and, in general, engagement with one's health and well-being. However, this study and other research have shown that certain demographics are more likely to use it than are others. In our study, younger females without disabilities and with higher income had the highest odds of mHealth use. As we are in the process of developing an app-based intervention for CVD and T2DM prevention, we are interested in finding strategies that would facilitate optimal uptake by those who would be most likely to benefit from the intervention. Our findings indicate that health-related apps were less used by people with a physical disability. We think that people with disabilities would likely benefit from more personalized interventions that take into account their disability. Moreover, the results showed that it will be important to consider equity issues to ensure that people with low income and older people will not be left out. Other researchers have drawn similar conclusions and proposed solutions. For example, Foley et al [41] used a mixed methods study to explore the access, use, and benefits of digital health services in Australia and identified trust as a key factor for digital health services use. They explained that recommendations by health professionals improved trust in digital health services, which could lead to increased uptake [41]. Cheng et al [42] concluded in their systematic review that the level of electronic health literacy was often overlooked when designing such interventions, leading to a digital divide with socially disadvantaged groups being left behind. Therefore, they demanded that electronic health literacy levels need to be recognized when developing interventions [42]. In their review, Borg et al [43] identified attitudes, skills, and access as barriers to digital inclusion, and social support, education, and inclusive design as enablers. The authors highlighted the importance of user-focused and collaborative designs to ensure digital inclusion [43].

Conclusions

Despite most people at risk of CVD or T2DM owning a smartphone, only about a third had ever used mHealth apps. There was no difference in mHealth use between people at risk of CVD or T2DM and those not at risk. People who used mHealth apps were less likely to be male, older, with physical disability, and with a lower income. This shows that it is important to consider equity issues when implementing a mHealth intervention. For example, low income or older age should not prevent people from participating in the intervention, and, therefore, these factors should be considered when developing an implementation strategy.

Acknowledgments

VHB was supported by a joint stipend from the University of New South Wales and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

This research was completed using data collected through the 45 and Up Study (The Sax Institute). The 45 and Up Study is managed by the Sax Institute in collaboration with the major partner, Cancer Council NSW; and other partners, including the Heart Foundation, the NSW Ministry of Health, the NSW Department of Communities and Justice, and Australian Red Cross Lifeblood. We thank the many thousands of people participating in the 45 and Up Study. Further, we would like to acknowledge Services Australia for supplying the PBS data used in this research project and SURE for the provision of secure data access.

Conflicts of Interest

None declared.

References

1. Bertram MY, Sweeny K, Lauer JA, Chisholm D, Sheehan P, Rasmussen B, et al. Investing in non-communicable diseases: an estimation of the return on investment for prevention and treatment services. Lancet 2018 Dec 19;391(10134):2071-2078. [doi: 10.1016/S0140-6736(18)30665-2] [Medline: 29627159]
2. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1223-1249 [FREE Full text] [doi: 10.1016/S0140-6736(20)30752-2] [Medline: 33069327]
3. Atun R, Jaffar S, Nishtar S, Knaul FM, Barreto ML, Nyirenda M, et al. Improving responsiveness of health systems to non-communicable diseases. Lancet 2013 Feb 23;381(9867):690-697. [doi: 10.1016/S0140-6736(13)60063-X] [Medline: 23410609]
4. Digital health trends 2021: Innovation, evidence, regulation, and adoption. IQVIA Institute for Human Data Science. 2021. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/digital-health-trends-2021/iqvia-institute-digital-health-trends-2021.pdf?_=1635260142390 [accessed 2022-02-07]
5. Jensen M, Treskes R, Caiani E, Casado-Arroyo R, Cowie M, Dilaveris P, European Association of Preventive Cardiology, European Heart Rhythm Association, Association of Cardiovascular Nursing and Allied Professionals. ESC working group on e-cardiology position paper: Use of commercially available wearable technology for heart rate and activity tracking in primary and secondary cardiovascular prevention, in collaboration with the European Heart Rhythm Association, European Association of Preventive Cardiology, Association of Cardiovascular Nursing and Allied Professionals, Patient Forum, and the Digital Health Committee. Eur Heart J Digit Health 2021;2(1):49-59. [doi: 10.1093/ehjdh/ztab011]
6. Meskó B, Drobni Z, Bényei, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38. [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
7. Spoth R, Rohrbach LA, Greenberg M, Leaf P, Brown CH, Fagan A, Society for Prevention Research Type 2 Translational Task Force. Addressing core challenges for the next generation of type 2 translation research and systems: the translation science to population impact (TSci Impact) framework. Prev Sci 2013 Aug;14(4):319-351 [FREE Full text] [doi: 10.1007/s11121-012-0362-6] [Medline: 23430579]
8. Shan R, Ding J, Plante TB, Martin SS. Mobile health access and use among individuals with or at risk for cardiovascular disease: 2018 Health Information National Trends Survey (HINTS). J Am Heart Assoc 2019 Dec 17;8(24):e014390. [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
9. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
10. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US population by patterns of mobile health use for health and behavioral tracking: analysis of the National Cancer Institute's health information national trends survey data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]
11. 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]
12. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 02;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]
13. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Dec 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
14. 45 and Up Study Collaborators, Banks E, Redman S, Jorm L, Armstrong B, Bauman A, et al. Cohort profile: the 45 and up study. Int J Epidemiol 2008 Oct;37(5):941-947 [FREE Full text] [doi: 10.1093/ije/dym184] [Medline: 17881411]

15. Bleicher K, Summerhayes R, Baynes S, Swarbrick M, Navin Cristina T, Luc H, et al. Cohort profile update: the 45 and Up Study. Int J Epidemiol 2022 May 23:dyac104. [doi: 10.1093/ije/dyac104] [Medline: 35604374]
16. Sax Institute. Use the 45 and Up Study. Sydney, Australia: Sax Institute; 2021. URL: https://www.saxinstitute.org.au/our-work/45-up-study/for-researchers/ [accessed 2022-02-07]
17. Questionnaires. Sax Institute. Sydney, Australia; 2021. URL: https://www.saxinstitute.org.au/our-work/45-up-study/questionnaires/ [accessed 2022-02-07]
18. Centre for Health Record Linkage. CHeReL. 2021. URL: https://www.cherel.org.au/ [accessed 2022-02-07]
19. Joshy G, Korda RJ, Attia J, Liu B, Bauman AE, Banks E. Body mass index and incident hospitalisation for cardiovascular disease in 158 546 participants from the 45 and Up Study. Int J Obes 2013 Oct 23;38(6):848-856. [doi: 10.1038/ijo.2013.192] [Medline: 24149770]
20. Comino EJ, Tran DT, Haas M, Flack J, Jalaludin B, Jorm L, et al. Validating self-report of diabetes use by participants in the 45 and Up Study: a record linkage study. BMC Health Serv Res 2013 Nov 19;13:481 [FREE Full text] [doi: 10.1186/1472-6963-13-481] [Medline: 24245780]
21. Australian Institute of Health and Welfare. Procedures data cube under ACHI Edition 9, 2016-17. Canberra, Australia; 2017. URL: https://www.aihw.gov.au/reports/hospitals/procedures-data-cubes/contents/data-cubes [accessed 2022-02-07]
22. Medicines for cardiovascular disease. Australian Institute of Health and Welfare. Canberra, Australia; 2017. URL: https://www.aihw.gov.au/getmedia/e84e445a-b4f0-4eac-96ee-b4cbf4e5639a/aihw-cvd-80.pdf.aspx?inline=true [accessed 2022-02-07]
23. Australian Institute of Health and Welfare. Heart, stroke and vascular diseases - Australian facts 2001. Canberra, Australia: AIHW National Heart Foundation National Stroke Foundation; 2001. URL: https://www.aihw.gov.au/getmedia/f49ee188-827d-4f12-ac78-25a7c823c98f/hsvd01.pdf.aspx?inline=true [accessed 2022-02-07]
24. Cardiovascular disease Australian facts 2011. Australian Institute of Health and Welfare. Canberra, Australia; 2011. URL: https://www.aihw.gov.au/getmedia/9621f6a8-f076-4e3e-a9c7-dece59ff0d74/12116.pdf.aspx?inline=true [accessed 2022-02-07]
25. Australian Government Department of Health. Schedule of Pharmaceutical Benefits: General pharmaceutical schedule - effective 1 September. Canberra, Australia; 2020. URL: https://www.pbs.gov.au/publication/schedule/2020/09/2020-09-01-general%20schedule-volume-1.pdf [accessed 2022-02-07]
26. The Pharmaceutical Benefits Scheme - Browse by body system. Australian Government Department of Health. Canberra, Australia; 2022. URL: https://www.pbs.gov.au/browse/body-system [accessed 2022-02-07]
27. Under pressure: The squeezed middle class. OECD. Paris, France; 2019. URL: https://www.oecd-ilibrary.org/content/publication/689afed1-en [accessed 2022-02-07]
28. Australian guidelines to reduce health risks from drinking alcohol. National Health and Medical Research Council. Canberra, Australia; 2020. URL: https://www.nhmrc.gov.au/about-us/publications/australian-guidelines-reduce-health-risks-drinking-alcohol [accessed 2022-02-07]
29. Australian dietary guidelines. National Health and Medical Research Council. Canberra, Australia; 2013. URL: https://www.health.gov.au/sites/default/files/australian-dietary-guidelines.pdf [accessed 2022-02-07]
30. Australia's physical activity and sedentary behaviour guidelines and the Australian 24-hour movement guidelines. Department of Health. Canberra, Australia; 2021. URL: https://www.health.gov.au/health-topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians?utm_source=health.gov.au&utm_medium=callout-auto-custom&utm_campaign=digital_transformation [accessed 2022-02-07]
31. SURE. Sax Institute. Sydney, Australia; 2021. URL: https://www.saxinstitute.org.au/our-work/sure/ [accessed 2022-02-07]
32. Seifert A, Vandelanotte C. The use of wearables and health apps and the willingness to share self-collected data among older adults. Aging and Health Research 2021 Sep;1(3):100032. [doi: 10.1016/j.ahr.2021.100032]
33. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
34. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology 2012 Sep;1(3):123-126. [doi: 10.1016/j.hlpt.2012.07.003]
35. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, China Kadoorie Biobank (CKB) collaborative group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011 Dec;40(6):1652-1666 [FREE Full text] [doi: 10.1093/ije/dyr120] [Medline: 22158673]
36. Scholtens S, Smidt N, Swertz MA, Bakker SJ, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015 Aug;44(4):1172-1180. [doi: 10.1093/ije/dyu229] [Medline: 25502107]
37. Mealing NM, Banks E, Jorm LR, Steel DG, Clements MS, Rogers KD. Investigation of relative risk estimates from studies of the same population with contrasting response rates and designs. BMC Med Res Methodol 2010 Apr 01;10:26 [FREE Full text] [doi: 10.1186/1471-2288-10-26] [Medline: 20356408]
38. Barr M, Baker D, Gorringe M, Fritsche L. NSW Population Health Survey: description of methods. New South Wales Department of Health. Sydney, Australia; 2008. URL: https://www.health.nsw.gov.au/surveys/adult/Documents/description-of-methods-2002-2011.pdf [accessed 2022-02-07]

39. Wang JJJ, Bartlett M, Ryan L. On the impact of nonresponse in logistic regression: application to the 45 and Up study. BMC Med Res Methodol 2017 May 08;17(1):80 [FREE Full text] [doi: 10.1186/s12874-017-0355-z] [Medline: 28482809]
40. Ng SP, Korda R, Clements M, Latz I, Bauman A, Bambrick H, et al. Validity of self-reported height and weight and derived body mass index in middle-aged and elderly individuals in Australia. Aust N Z J Public Health 2011 Dec;35(6):557-563. [doi: 10.1111/j.1753-6405.2011.00742.x] [Medline: 22151163]
41. Foley K, Freeman T, Ward P, Lawler A, Osborne R, Fisher M. Exploring access to, use of and benefits from population-oriented digital health services in Australia. Health Promot Int 2021 Aug 30;36(4):1105-1115. [doi: 10.1093/heapro/daaa145] [Medline: 33367568]
42. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
43. Borg K, Boulet M, Smith L, Bragge P. Digital inclusion & health communication: a rapid review of literature. Health Commun 2019 Oct;34(11):1320-1328. [doi: 10.1080/10410236.2018.1485077] [Medline: 29889560]

Abbreviations

AIHW: Australian Institute of Health and Welfare
APDC: Admitted Patient Data Collection
CSIRO: Commonwealth Scientific and Industrial Research Organisation
CVD: cardiovascular disease
ICD-10-AM: International Classification of Diseases version 10 Australian Modification
mHealth: mobile health
NSW: New South Wales
OECD: Organization for Economic Cooperation and Development
OR: odds ratio
PBS: Pharmaceutical Benefits Scheme
SURE: Secure Unified Research Environment
T2DM: type 2 diabetes mellitus

Edited by L Buis, G Eysenbach; submitted 18.02.22; peer-reviewed by W Jiang, Y Xie, R Ghosh; comments to author 22.06.22; revised version received 15.07.22; accepted 24.08.22; published 07.09.22.

Please cite as:
Buss VH, Varnfield M, Harris M, Barr M
Mobile Health Use by Older Individuals at Risk of Cardiovascular Disease and Type 2 Diabetes Mellitus in an Australian Cohort: Cross-sectional Survey Study
JMIR Mhealth Uhealth 2022;10(9):e37343
URL: https://mhealth.jmir.org/2022/9/e37343
doi:10.2196/37343
PMID:36069764

Corrigenda and Addenda

Authorship Correction: Impact of a Mobile Application for Tracking Nausea and Vomiting During Pregnancy (NVP) on NVP Symptoms, Quality of Life, and Decisional Conflict Regarding NVP Treatments: MinSafeStart Randomized Controlled Trial

Elin Ngo[1], MSc; Maria Bich-Thuy Truong[1], PhD; David Wright[2,3], PhD; Hedvig Nordeng[1,4], PhD

[1]PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo, Oslo, Norway

[2]School of Allied Health Professions, University of Leicester, England, United Kingdom

[3]Centre for Pharmacy, University of Bergen, Bergen, Norway

[4]Department of Child Health and Development, National Institute of Public Health, Oslo, Norway

Corresponding Author:
Elin Ngo, MSc
PharmacoEpidemiology and Drug Safety Research Group
Department of Pharmacy
University of Oslo
Postbox 1068 Blindern
Oslo, 0316
Norway
Phone: 47 93849866
Email: e.t.p.ngo@farmasi.uio.no

Related Article:

Correction of: https://mhealth.jmir.org/2022/7/e36226

(JMIR Mhealth Uhealth 2022;10(9):e41927) doi:10.2196/41927

In "Impact of a Mobile Application for Tracking Nausea and Vomiting During Pregnancy (NVP) On NVP Symptoms, Quality of Life, and Decisional Conflict Regarding NVP Treatments: MinSafeStart Randomized Controlled Trial" (JMIR Mhealth Uhealth 2022;10(7):e36226) the authors noted one error.

In the originally published article, authors Hedvig Nordeng and David Wright were ordered as, respectively, third and fourth authors. The correct order of authors should be David Wright as third author and then Hedvig Nordeng as fourth author. Author affiliations have been reordered accordingly.

The complete list of authors and affiliations in the originally published article was as follows:

Elin Ngo[1], MSc; Maria Bich-Thuy Truong[1], PhD; Hedvig Nordeng[1,2], PhD; David Wright[3,4], PhD

[1]PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo, Oslo, Norway

[2]Department of Child Health and Development, National Institute of Public Health, Oslo, Norway

[3]School of Allied Health Professions, University of Leicester, England, United Kingdom

[4]Centre for Pharmacy, University of Bergen, Bergen, Norway

The list of authors and affiliations has been corrected to:

Elin Ngo[1], MSc; Maria Bich-Thuy Truong[1], PhD; David Wright[2,3], PhD; Hedvig Nordeng[1,4], PhD

[1]PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo, Oslo, Norway

[2]School of Allied Health Professions, University of Leicester, England, United Kingdom

[3]Centre for Pharmacy, University of Bergen, Bergen, Norway

[4]Department of Child Health and Development, National Institute of Public Health, Oslo, Norway

The correction will appear in the online version of the paper on the JMIR Publications website on September 26, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.08.22; this is a non–peer-reviewed article;accepted 18.08.22; published 26.09.22.

Please cite as:
Ngo E, Truong MBT, Wright D, Nordeng H
Authorship Correction: Impact of a Mobile Application for Tracking Nausea and Vomiting During Pregnancy (NVP) on NVP Symptoms, Quality of Life, and Decisional Conflict Regarding NVP Treatments: MinSafeStart Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(9):e41927
URL: https://mhealth.jmir.org/2022/9/e41927
doi:10.2196/41927
PMID:36162103

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org